

nature



ANXIETY DISORDERS

Compulsive behaviour
in a mouse model

OLD AS THE HILLS

Ancient diamonds
near to history

HOMED DEIGNS

Exhibit the gorilla

NEURAL MIGRATION

Gap junctions stick to their task

TECHNOLOGY
FEATURE

Virtual neurons

Innovation versus science?

Harder economic times will force governments to ask tough questions about their investments in research.

Science and science funding have enjoyed a good run over the past 20 years or so. During the unusually long economic boom after the end of the cold war, governments of all political stripes have accepted the argument that it is in their interest to support 'the best science'. Something like a global consensus has emerged on the value of curiosity-driven, basic research. The predominant argument behind this consensus has been the belief that excellent science — of whatever discipline — is likely to spur innovation, which will in turn foster economic growth.

At the same time, direct government sponsorship of technology development has fallen out of vogue. Considerable expenditure continues, of course, and if the military sphere is included, it still dwarfs the resources devoted to basic science. But the idea of explicit state support for the development of drugs or circuit boards or civilian airliners has been pretty well driven off the table. There has been an assumption — even in comparatively centrist nations such as France and Japan — that governments aren't good at 'picking winners'.

This particular era of science and technology policy may now be drawing to a close. Even before last week's juddering stock markets sparked talk about a possible US, or even global, recession, policymakers in industrialized countries were watching the flight of much industrial production to China and India with intensifying alarm.

As a result, the word 'competitiveness' is back on the agenda, particularly in the United States. When political leaders look at research budgets in the light of competitiveness, they are always prone to be more drawn to the direct support of innovation through technology programmes, than to its indirect support, through basic science. As a result, they may start asking questions that are inherently difficult for scientists to answer, such as, what will be the economic spin-off from this work? What are we getting for our money?

These questions are currently being posed most directly in the United Kingdom, where the research councils, which support most university science, seem to be undergoing a subtle change of direction. Some scientist groups are already nervous about a paper, "Increasing the economic impact of the research councils", that was published in January by the councils' steering group.

There is a risk that this process could result in perceived economic

relevance displacing scientific merit, to a significant extent, as the determining factor in the selection of research-council grants. Before that happens, it would be reassuring if the leaders of the research councils emphatically reiterated that their primary function is to promote scientific excellence — and that they will best support Britain's national interests, economic and otherwise, by doing exactly that.

A similar trend is beginning to emerge in the United States. The competitiveness legislation passed by Congress and signed into law on 9 August is positive for science — it lays heavy emphasis on supporting basic research in mathematics, the physical sciences and engineering. Yet it is by no means certain that this approach will be reflected in the annual budget process. The Democrats, who now control both houses of Congress, have traditionally favoured technology programmes, sometimes over scientific ones, and that pattern could resume as fears about national competitiveness become more acute.

The argument will then be made for different scientific programmes to 'prove themselves' in answer to the taxpayers' question: what have you done for us lately? It is important, in such an environment, for scientists to hold their nerve. In particular, they should resist the inevitable demand that they start to compete with each other on the basis of specious metrics that bureaucrats, given half a chance, will construct as surrogates for economic impact.

Straitened economic circumstances, whenever they arrive, will mean straitened times for science: that much the research community must be ready to accept. It can also anticipate attempts to cut and splice the pie on the basis of ungrounded expectations that some disciplines will yield economic dividends, and others won't. Researchers have at least two weapons that they should keep well honed: a compelling historical narrative showing the unpredictable paths from science across all disciplines to economic and other benefits; and a demonstration that those best-placed to innovate on the basis of science — and, in turn, to stimulate scientific ideas — are well set up to do so. ■

"Scientists should resist the demand to compete on the basis of specious metrics constructed as surrogates for economic impact."

Indentured labour

The deal at the foot of the scientific totem pole remains a raw one.

The Federation of American Societies for Experimental Biology (FASEB) has just released a report on the career trajectories of young life scientists in the United States (see page 848). It is likely to give pause to some of those currently considering graduate

training as a route to a career in the academic life sciences.

The survey finds that over two decades the number of academically employed life scientists in tenured or tenure-track positions has remained stuck at about 30,000, while the number of doctoral degrees awarded in the life sciences has doubled. Thus the proportion of postdocs actually reaching tenured or tenure-track positions has dropped from nearly 45% in the early 1980s, to just below 30%.

The data also reveal a hard-to-reach career getting farther out of reach. The age at which the average PhD holder receives his or her first full National Institutes of Health grant has risen from 34 in 1970

to 42 now. Postdocs, facing such a late start to their professional lives, are increasingly jumping ship to industry.

Academic institutions rely heavily on graduate students and postdocs to bring in tuition or overhead funding and to carry their share of the teaching load. The motivation for principal investigators is even stronger. Students and postdocs carry out the day-to-day work in laboratories serving as cheap, well trained labour. Moreover the nature of discovery often seems to require big numbers: far better to have six postdocs working on several projects, in case one of them gets results that will ensure funding for the laboratory for years.

This pattern has, of course, been familiar for years — and not just in the United States. Postdocs find themselves bouncing around the world from lab to lab, seldom earning much more than they would have done in their first year on the job market with their undergraduate degree. Funding is short, the hours are long, and prospects uncertain.

Postdocs have occasionally attempted to band together in solidarity and seek a better settlement from their employers, the institutions and universities. But this movement has been stronger in the social sciences than in the hard sciences. The transient nature of the

work, together with its convoluted employment structure, has made it difficult for them to speak effectively with a single voice. Instead, the plight of the postdoc will probably change only if the issue of scientific training is addressed from the top, where it may be necessary to consider the possibility that too many scientists are being trained.

There is an argument that, from a national policy perspective, the current situation is ultimately productive. The pace of discovery is quickened by a sizeable workforce, and able scientists end up doing multiple jobs, most of them in the private sector of the economy. It might not be exactly what the students had in mind in the first place, but the situation hardly constitutes a major cause for concern.

But FASEB's data suggest that too many graduate schools may be preparing too many students, so that too few young scientists have a real prospect of making a career in academic science. More effort is needed to ensure that recruitment interviews include realistic assessments of prospective students' expectations and potential in the academic workplace. And training should address broader career options from day one rather than focusing unrealistically on jobs that don't exist. ■

"Funding is short, the hours are long, and prospects uncertain."

Technology trap

California is right to sound a cautionary note on electronic voting.

Designing an electronic voting system that is easy to use, efficient and secure may sound like an easy thing to do. And the pay-off — a democracy in which more people can participate and trust — seems desirable. But an academic analysis of three widely used systems in California has found monumental weaknesses in each of them. As a result, the state is slowing down its adoption of such systems until significant improvements are made. Others should exercise similar caution.

The study, commissioned by California's secretary of state, Debra Bowen, was led by computer scientists at the Berkeley and Davis campuses of the University of California. It found that the systems sold by three companies — Sequoia Voting Systems, Hart InterCivic and Diebold — had not been designed with security requirements in mind. And one particular deficit alarmed representatives of all political parties: the possibility that computer viruses could distort vote counts.

On 3 August, Bowen decertified the systems, which were already in use in counties where about half of the state's voters live. That means that in the primary elections next February, voters will return to paper ballots. Bowen has pledged to fully recertify the machines when they comply with a list of basic requirements: but the study authors question whether the software and hardware are amenable to ready repair. "They have serious security problems that will take years to fix," says David Wagner, a study leader at the University of California, Berkeley.

This isn't the first time that specialists have warned against electronic voting systems. The Voting Technology Project, for example, a joint effort between the Massachusetts and California Institutes

of Technology, highlighted their failings back in 2001 (see *Nature* **412**, 258; 2001).

Yet the march of voting automation continues worldwide, often driven not by the public good but by election officials' desire for low staff costs and quick counts — as well as by the marketing machines of the systems' suppliers. Even in the United States, the Californian analysis is unlikely to make much of a difference in the many other states where the same electronic systems are being introduced. Verifiedvoting.org, a non-partisan lobby group that campaigns for reliable voting, says that although some secretaries of state are paying attention to the study, others — especially in the south and the midwest — don't seem to be interested.

There remains a body of public officials who seem to favour expediency and convenience over the democratic imperative of an accurate count. The firms that sell the systems have, meanwhile, argued that in the real world of elections, the systems will be overseen by election officials and candidates who would protect against the kind of disruptions identified in laboratory studies.

After the scandal that unfolded in Florida in the 2000 presidential election, when President George W. Bush eked out a narrow victory after prolonged legal arguments over disputed ballots in several counties, Congress passed a law that, among other things, helps to fund the replacement of existing, outmoded voting equipment. Now it is set to revisit the issue, with Senator Dianne Feinstein (Democrat, California) pledging to hold hearings that will pick up where the review in her own state left off. This may spur broader federal action to strengthen voting systems.

The consistent message from studies of electronic voting systems has been that the technology is often being implemented before it is ready to achieve the levels of security and reliability that voters are entitled to expect. Other jurisdictions worldwide should follow California's lead, consult with computer scientists, and act where necessary to stop this from happening. ■

RESEARCH HIGHLIGHTS

The shadow of drought

Geophys. Res. Lett. **34**, L15702 (2007)

The worldwide cloud of aerosols produced by the eruption of Mount Pinatubo in 1991 (pictured) led to a remarkable slowing of Earth's hydrological cycle, according to researchers at the National Center for Atmospheric Research in Boulder, Colorado.

Kevin Trenberth and Aiguo Dai show that 1992 was deficient in both precipitation over land and streamflow to a degree unseen in 55 years of records, probably as a result of reduced incoming solar radiation leading to reduced evaporation. The effect remained significant even when the impact of a near-concurrent El Niño event was factored out.

The authors warn that 'geoengineering' projects designed to offset greenhouse warming by reducing the amount of incoming solar radiation on a global scale may have similar consequences.



D. HARLOW/US GEOLOGICAL SURVEY/TIME LIFE PICTURES/GETTY

CELL BIOLOGY

Little helper

Science **317**, 957–961 and 961–964 (2007)

Two studies offer molecular insight into how proteins get into and across cellular membranes.

Both studies suggest that protein structures known as POTRA domains, found in the outer membranes of Gram-negative bacteria, mitochondria and chloroplasts, begin the job by binding membrane-destined proteins.

Vincent Villeret at the Institute of Biology of Lille, France, Françoise Jacob-Dubuisson at the Lille Pasteur Institute and their co-workers solved the structure of part of a membrane protein known as FhaC from *Bordetella pertussis*, which causes whooping cough. FhaC forms a channel that transports virulence proteins across the membrane, opening when a portion of one of its POTRA domains binds the virulence protein.

Daniel Kahne at Harvard University in Cambridge, Massachusetts, and his colleagues studied a fragment of a protein known as YaeT from *Escherichia coli*. It forms a scaffold for a complex that folds other membrane proteins. A portion of one of its POTRA domains seems to bind these proteins.

CHEMISTRY

Blinking hot

J. Am. Chem. Soc. doi:10.1021/ja0715905 (2007)

You can't use a medicine-cabinet thermometer in a microlitre of fluid. Scientists wanting to take the temperature of reagents in microfluidic devices might turn to a 'blinking' molecule instead.

The proposal comes from Cécile Fradin and her team at McMaster University in Hamilton, Canada. They note that under laser illumination, enhanced green fluorescent protein switches between a glowing and a

dull state with the removal and addition of a proton. How fast the protein 'blinks' depends on two properties of the solution to which it is added: acidity and temperature.

Previous methods to gauge temperature with dyes have relied on changes in fluorescence intensity, which yield only relative measurements. This protein, at known pH, gives an absolute temperature measurement.

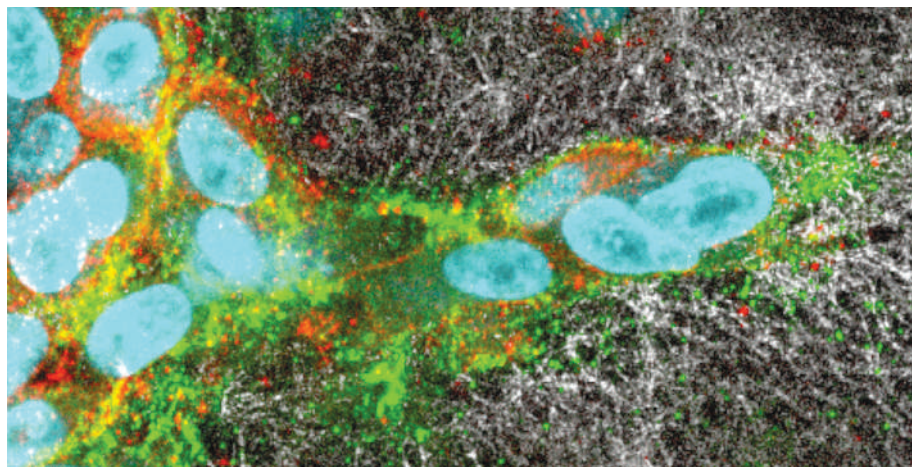
CANCER BIOLOGY

Tube travel

Nature Cell Biol. **9**, 893–904 (2007)

With the help of time-lapse microscopy, researchers have revealed how tumours degrade and invade the protein matrix around healthy cells to spread through the body.

Peter Friedl at the University of Würzburg in Germany and his colleagues studied fibrosarcoma and breast cancer cells as the cells invaded three-dimensional blocks of collagen in a lab dish. The team identified two key steps. First, single cells rearrange collagen fibres that are blocking their way into tube-like holes. Then, groups of cells behind the leading cells expand the tubes (pictured left), clearing away larger areas of the matrix with the help of enzymes including membrane-type-1 matrix metalloproteinase (MT1-MMP).



SOLID-STATE PHYSICS

All for one

Nano Lett. **7**, 2506–2512 (2007)

In today's silicon solar cells, which account for more than 90% of the solar-cell market, each absorbed photon produces a pair of charge

carriers of fixed energy. Any surplus energy from the original photon is lost as heat.

However, in nanoparticles of certain semiconductors, such as lead selenide, a single photon that has enough energy to produce two or more pairs of charge carriers can do so. Now Matt Beard, Randy Ellingson, Arthur Nozik and other colleagues at the National Renewable Energy Laboratory in Golden, Colorado, show that the same can happen in silicon nanoparticles. This 'multiple exciton generation' would give silicon-nanoparticle-based cells a theoretical peak efficiency of about 44%, compared with around 33% for traditional silicon solar cells. But the researchers have yet to work out how to make use of the charge carriers in such particles.

GENETICS

Red or white

PloS Genet. **3**, e133 (2007)

Researchers have surveyed the genomes of 440 rice strains to work out the evolutionary history of how wild rice lost its red colour to cultivation.

Susan McCouch at Cornell University in Ithaca, New York, and her colleagues report that more than 97% of the white-grained rices they studied have the 14 base-pair deletion previously identified with the loss of the red outer layer, or pericarp, of wild rice grains. Other clues in the genome suggest this mutation arose once in *japonica* subspecies and then spread around the world, probably through human trade. It is thought that humans might have preferred white rice over red for its novelty, its softer shell, or because it was easier to pick out insects or dirt from a meal. A second mutation associated with white grains appeared in fewer than 3% of species studied.

ECOLOGY

Worm wood

Conserv. Biol. **21**, 997–1008 (2007)

Traipsing through the sugar-maple forests of North America has become a different experience since alien earthworms invaded.

Andrew Holdsworth and his team at the University of Minnesota in St Paul have examined various sites in two National Forests in the northern United States. They found a clear negative correlation between the presence of worms from the European genus *Lumbricus* (pictured right) and plant diversity, particularly the diversity of small non-grassy plants, or forbs.

This confirms past observations of *Lumbricus* disrupting native plant assemblages. The suggested reason? The European worms eat leaf litter that would otherwise shelter seeds and host nutrient-providing fungi. As a result, species such as wild sarsaparilla (*Aralia nudicaulis* L.) and rose twisted stalk (*Streptopus roseus* Michx.) disappear.

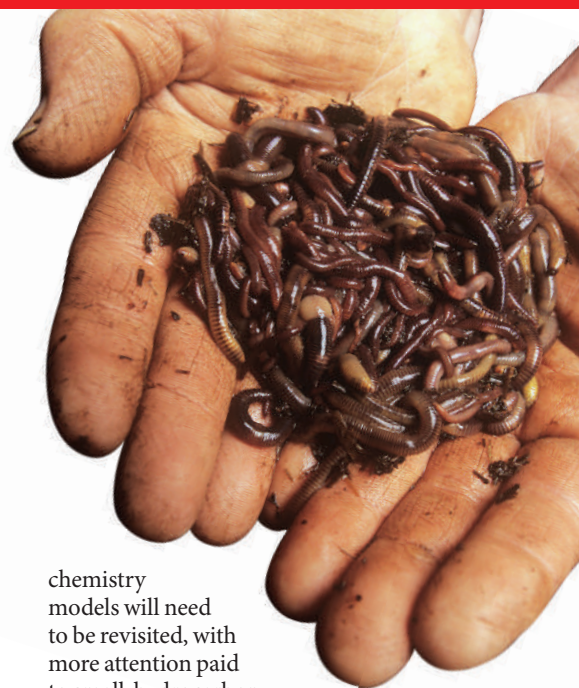
ASTRONOMY

Chemical misfit

Astrophys. J. **665**, L127–L130 (2007)

Propene has been detected in space for the first time, in a dark interstellar cloud in the constellation of Taurus. No one had previously looked for the molecule (C_3H_6) because, although such clouds can be rich in hydrocarbons, there seemed to be no obvious mechanism by which propene would form.

Núria Marcelino at Instituto de Estructura de la Materia in Madrid, Spain, and her colleagues made the discovery with the IRAM 30-metre radio telescope near Granada. Marcalino's team suggests that now they have found propene, interstellar



chemistry models will need to be revisited, with more attention paid to small-hydrocarbon chemistry.

QUANTUM PHYSICS

Massive interference

Nature Phys. doi:10.1038/nphys701 (2007)

Shooting a beam of electrons, atoms or molecules through a grating is the classic way to reveal the particles' quantum wave-like nature. The beam produces an interference pattern, just as light would.

Markus Arndt of the University of Vienna, Austria, and his colleagues report a nifty way to extend this technique to larger molecules than it has been possible to study so far.

Large molecules tend to interact with a grating, blurring the interference pattern. Arndt's team overcome this problem with a clever arrangement of three gratings — the second made not of material, but of a 'standing wave' of light. The properties of the light barrier can be tuned to optimize the contrast of the interference image, as the authors demonstrate for ^{70}C — a fullerene — and azobenzene molecules.

S. STAMMERS/SPL

JOURNAL CLUB

David Stevenson
California Institute of
Technology, USA

A planetary scientist foresees a shift in the debate about Earth's heat flow.

Measurements of the heat coming out of Earth's interior have long posed a puzzle for understanding the planet's history.

Earth's heat output is estimated to be around 44 terawatts, about twice that expected from radioactive decay. The difference

can be attributed to cooling of the deep Earth, implying a present-day cooling rate of 100 kelvin per billion years. But simple models with this much cooling 'blow up' when they are run back in time, predicting ridiculous temperatures for the early Earth. Acceptable models rely on unconventional deviations from the usual simple scaling laws for mantle convection. This is an attractive but untested idea.

I and many others have wondered whether an alternative explanation is that today's heat flow is higher than the average for

the past half a billion years. Such fluctuations could arise as a result of the dispersal and accumulation of continental land masses.

A recent paper (J. Korenaga *Earth Planet. Sci. Lett.* **257**, 350–358; 2007) assessed this possibility by taking advantage of a long-known connection between sea level and the heat flow from sea-floor spreading. It finds little room for more than a few percent fluctuation in heat flow around its long-term decline.

I think this pushes the problem back into the realm of models, focusing attention on plate

tectonics, the deep water cycle (because water affects how rocks flow), and perhaps even the long-standing question of whether Earth's mantle is well mixed from top to bottom.

On a decadal timescale, we can hope that better measurements of heat generation and flow will be combined with more realistic theory. Like many central Earth science questions, the heat-flow problem resists quick resolution.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS

Oldest gorilla ages our joint ancestor

The last common ancestor of humans and gorillas might have lived at least 2 million years earlier than previously thought. Fossilized teeth of the earliest gorilla ever discovered, dating to 10 million years ago, have been found in Africa, say researchers.

The new species (*Chororapithecus abyssinicus*) from Ethiopia, reported on page 921 of this issue, helps to fill in a huge gap in the fossil record. The team of Ethiopian and Japanese researchers has based its conclusion on just nine teeth from at least three individuals of the species, which were discovered in the desert scrubland of Afar about 170 kilometres east of Addis Ababa.

The teeth, eight molars and a canine, “are collectively indistinguishable from modern gorilla subspecies” in size, proportion and scan-revealed internal structure, says Gen Suwa of the University of Tokyo Museum, Japan, who led the study. The team argues that the gorilla’s divergence date from the human lineage is not about 8 million years ago as previously surmised (S. Kumar *et al. Proc. Natl Acad. Sci. USA* **102**, 18842–18847; 2005), but “greater than 10 to 11 million years ago” on the basis of the age of

the new species. Functionally, he adds, the teeth already seem to be evolving — they could shear through a plant diet, a gorilla trait — although other herbivore apes also exist in the fossil record.

This finding could prompt discussions of how anthropologists and geneticists determine the hominin line’s divergence from chimps, previously pegged at about 6 million years ago. “*Chororapithecus abyssinicus* suggests, once again, that Africa was the place of origin of both humans and modern African apes” — not Eurasia as some researchers have argued, says Suwa.

But palaeoanthropologist Jay Kelley, who studies primate teeth at the University of Illinois at Chicago and was not involved in the study, is sceptical. “I’m not convinced it is a gorilla,” he says. More fossils, analysis and debate will be needed to determine whether



"Africa was the place of origin of both humans and modern African apes."

the specimen is ancestral to hominids, he adds. For now, he would be “very cautious” about using the specimen to realign divergence dates between hominins and gorillas–chimps.

Suwa’s team is part of the Revealing Human Origins Initiative (RHIO), a project that searches at multiple sites in Africa, Europe

Ocean circulation noisy, not stalling

Suggestions of a substantial decline in the Atlantic Ocean circulation that carries warm tropical water northwards seem to be largely unfounded. New data — recorded between the Bahamas and the Canary Islands — show that the seemingly dramatic reduction discovered two years ago in the strength of the Atlantic meridional overturning circulation (MOC) is easily within the range of huge seasonal variability.

In 2005, a team led by Harry Bryden, an oceanographer at the University of Southampton, UK, reported a 30% decrease

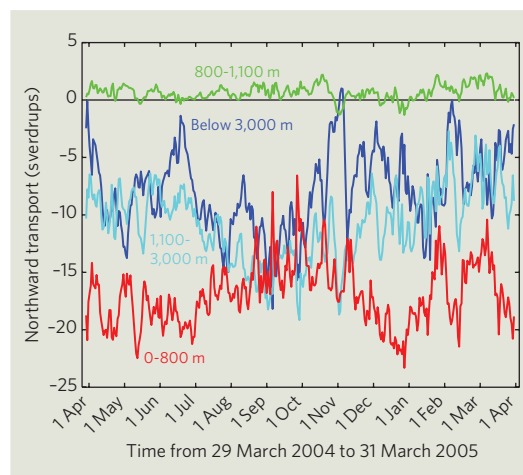
in the MOC at 26.5° north, after comparing their measurements from ship-based instruments in 2004 with similar data from 1957, 1981, 1992 and 1998 (H. L. Bryden, H. R. Longworth and S. A. Cunningham *Nature* **438**, 655–657; 2005). They did caution at the time that, due to the scarcity of observations, the error bounds of their study were exceptionally wide.

Nevertheless, the finding fuelled fears of an imminent collapse of the thermohaline Atlantic circulation — the ocean conveyor driven by temperature and salinity differences that gives

rise to the Gulf Stream, allowing western Europe to enjoy a relatively mild climate. Alarmist headlines warned that a mini ice age would hit Britain.

However, 12 months’ worth of data from an array of moored instruments, deployed during the same 2004 cruise that yielded the most recent measurements included in Bryden’s study, now suggest that the observed changes were due to short-term variability and not the result of global warming.

Stuart Cunningham of the National Oceanography Centre in Southampton and his colleagues found that the



During just one year, Stuart Cunningham’s team found enormous variability in the northwards flow of thermocline layer circulation (red), the deep waters of the upper (light blue), and the lower (dark blue) North Atlantic. Even the flow of intermediate water (green) fluctuated.



Fossilized teeth from the earliest-found gorilla (top, left), discovered at a site in Ethiopia (above), are remarkably similar to those of a modern gorilla (bottom, left).

G. SUR

and Asia for species that predate the earliest known hominid, the 7-million-year-old *Sahelanthropus tchadensis* (M. Brunet *et al. Nature* **418**, 145–151; 2002).

Between 15 million and 22 million years ago, there were dozens of primate species across Africa and Euroasia — and apes dominated the primate scene. But fossils show that these species don't share the characteristics

of modern African apes. "From that species pool, the common ancestor of African apes and humans branched out," says Tim White, a palaeoanthropologist at the University of California, Berkeley, and an RHOI director. "And the goal of RHOI is to find the common ancestors. With *C. abyssinicus*, we now can see an ancestral African ape."

Rex Dalton

strength of the overturning circulation varied wildly — by a factor of 8 — between March 2004 and March 2005 (S. A. Cunningham *et al. Science* **317**, 935–937; 2007). Bryden, who was also involved in the new study, had unknowingly probed the ocean during a period of relatively low circulation, they conclude.

"I guess we can give [thermohaline] the all-clear," says Jochem Marotzke, an oceanographer at the Max Planck Institute for Meteorology in Hamburg, Germany, formerly principal investigator of Rapid Climate Change (RAPID), a monitoring programme funded by the UK Natural Environment Research Council. The Atlantic sensor array, which is to remain operational until 2014, is part of the RAPID programme.

Available data provide no statistically significant evidence for a change of the overturning circulation over the past 50 years, says Marotzke, a co-author on the new study.

From now, the RAPID system should allow researchers to detect circulation changes from one year to the next, provided the average circulation volume changes by 20% or more. But because the unknown variability from year to year may also be large, it will take at least ten more years of continuous measurements until a possible downward trend will become recognizable in the data noise.

But Bryden remains unconvinced that the RAPID measurements in the new study have rendered his

initial findings null and void. The depth of the thermocline — the layer of rapid heat transfer between the mixed-water layer at the surface and deep-ocean water — varies across the basin from west to mid-ocean, changing the recirculation of water. An increase in the slope of the thermocline along with the observed freshening of deep water near the Bahamas, do indicate a 10–15% decline in average circulation since 1980, Bryden maintains.

"RAPID is quite an achievement," Bryden says. "But I don't share the opinion that our previous results are just due to seasonal variability. It may not be statistically reliable, but all the science points to a slow down of the Atlantic circulation." ■
Quirin Schiermeier

ON THE RECORD

"In our courses you will have the possibility to renew the natural contact you have with the angels. Use angels and your own power to create miracles in your own life."

Norway's Princess Märtha Louise's school for communicating with angels opens on 22 August and courses are already sold out. But TV pastor Jan Hanvold has unkindly called the clairvoyant princess "an emissary from hell". Sidelines wonders who the angels prefer to talk to.

SCORECARD

Left-clawed crabs
You'd think that being a left-clawed Australian fiddler crab when 98% are right-clawed would give you an advantage in a fight. But it turns out that left-clawed crabs are less likely to fight, and when they do, they usually lose. Perhaps that's why they are so much rarer...



Ghost crabs
In a study involving off-road vehicles on sandy Australian beaches scientists discovered that no ghost crabs buried less than 5 cm deep survive being run over 10 times, while more than 70% survive if buried at depths of 20 cm. Someone should tell the crabs.

M. JENNIONS

NUMBER CRUNCH

26 is the latest estimate of the maximum number of moves needed to solve any Rubik's cube.

63 hours is the time a supercomputer took to whittle down to this number from a best estimate of 29 moves.

20-something is the theoretical estimate of the maximum number of moves ever needed — a number still unproven.

43 billion billion is the total number of Rubik's cube positions.

Sources: BBC, Spiegel, Proc. R. Soc. B, Mar. Ecol.

Fleece myth hints at golden age for Georgia

THE ART ARCHIVE/CORBIS

The legend of Jason and the Argonauts' heroic quest for the Golden Fleece may provide a fascinating window into real Bronze Age society, say researchers who believe they have found the earliest evidence of a trading gold mine.

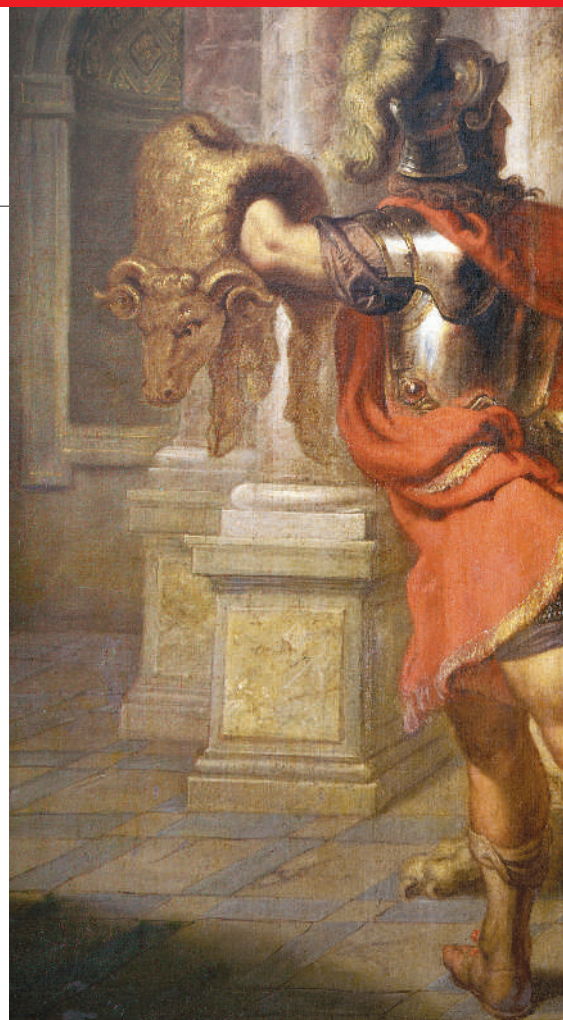
The adventure, which was set about 1300 BC — a generation before the Trojan war described by Homer — involves a mission to capture the precious fleece from Colchis, in modern-day Georgia. The mythological fleece has been interpreted in various ways, but is thought by some to symbolize the vast wealth and gold of Georgia.

Although gold was used and traded in ancient times, there has been scant evidence of where the precious metal used in Eurasia during the Bronze Age originated, and whether it was mined or panned from rivers. Gold mines in the Upper Nile region of Egypt and elsewhere in Africa date from 2,500 BC and are believed to have supplied the Ancient Egyptians' needs. But the legend of the Golden Fleece hints at a thriving gold trade in

the Caucasus dating back further than was first thought.

In 2004, a team of German and Georgian archaeologists discovered what may be the world's oldest gold mine at Sakdrissi, about 70 kilometres southwest of Tbilisi. Thomas Stöllner and Andreas Hauptmann of Ruhr University in Bochum, Germany, and Irina Gambashidze, an archaeologist from the Centre of Archaeological Studies in Tbilisi, dated the mine to around 3000 BC.

Now, the researchers have produced evidence that suggests this mine was trading as far back as 2500 BC. The team analysed samples of gold from the mine and compared them with gold from some 25 ancient artefacts dating from 2500 to 1500 BC in the Bronze Age. The beautifully worked pieces, housed in museum collections in Georgia and neighbouring countries, include needles, necklaces and rings. The gold was analysed using a process called LA-ICP-MS (laser ablation-inductively coupled plasma-mass



Analysis of Bronze Age artefacts, such as the lion (left), offers historical links to the Golden Fleece.

spectrometry), in which a tiny layer of surface metal is vaporized and the various elements in the vapour are identified by their mass.

Journal presents a mathematical conundrum

A branch of mathematics known as K-theory has this month seen its flagship journal bifurcate in a strange dispute, the details of which were still emerging as *Nature* went to press.

The events centre on Anthony Bak, a mathematician at the University of Bielefeld in Germany and until recently managing editor of the Springer journal *K-Theory*. Bak left the monthly journal, due, he says, to production problems and a too-high subscription rate. He has arranged to launch a rival title with another publisher that has offered a lower subscription rate, in the process shedding some editors unhappy at how the shift has been managed.

In seeking to cut prices, the editors follow the lead of the

editorial board of the Elsevier mathematics journal *Topology*, which quit to launch a competing non-profit title (see *Nature* 445, 351; 2007).

The cover date of the last issue of the supposedly monthly *K-Theory* was April 2006. Bak has withheld all papers accepted by the journal since then — a situation that not all of the editors, or the papers' authors, seem to have been aware of until recently.

In January 2007, the publisher sacked Bak. That month Bak wrote to his editorial board members asking them to resign en masse, which they did. As well as the cost of the journal, he mentioned production problems at Springer. But he neglected to tell Springer of the

resignations. "At no point did Bak communicate to Springer directly that the board had resigned," says Catriona Byrne, editorial director for mathematics at Springer.

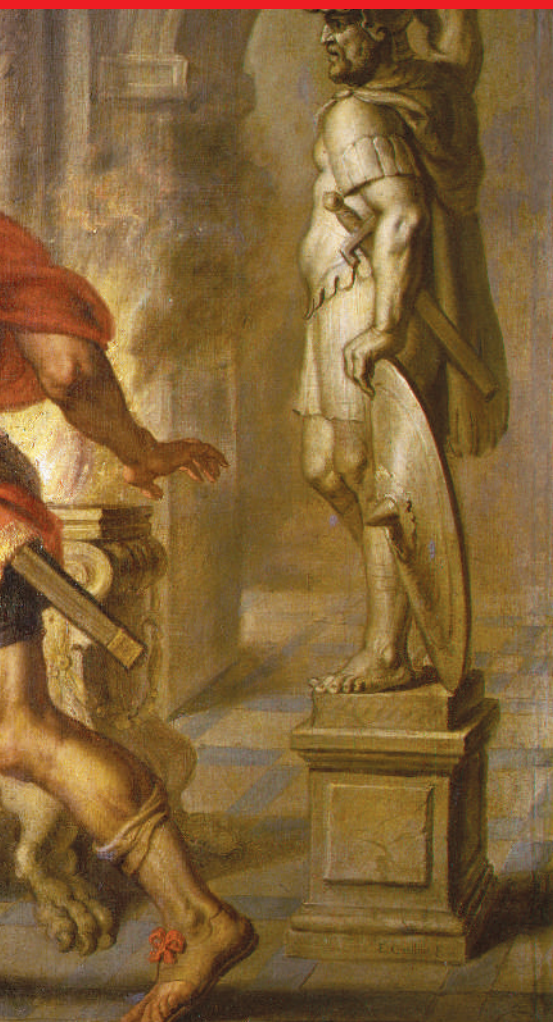
It was only in May that Springer learned that the board had resigned. In August, Bak announced publicly that he had arranged to launch a rival title, the *Journal of K-Theory*, with Cambridge University Press at a lower subscription rate. Springer has now said that it will lower the *K-Theory* subscription rate.

Eric Friedlander at Northwestern University in Evanston, Illinois, a former editor at *K-Theory*, is in principle sympathetic to the switch. "There is a lot of concern in the mathematics community about the cost of journals," he says. But on

17 August, Friedlander wrote to Bak to say that despite being named as a member of the editorial board of the new journal, he was not willing to serve "because I cannot endorse the process by which you have withheld manuscripts submitted to *K-Theory* and proceeded without consultation with authors and the editorial board".

One of the editors who resigned, Andrew Ranicki of the University of Edinburgh, UK, has returned to the Springer journal. He and mathematician Wolfgang Lück of the University of Münster, Germany, issued a statement on 13 August to say that they would act as interim managing editors for the journal.

Bak launched *K-Theory* in the 1980s and ran it successfully



The ratio of trace elements in the gold helps to pinpoint where it came from. In this case, the team found similar levels of lead and osmium in the gold from the mine in Sakdrissi as in the artefacts. But the clincher was the unusually high amount of antimony in the artefacts,

which is characteristic of the particular geochemistry at Sakdrissi.

Although this suggests a strong link between the mine and the artefacts, it is possible that the gold used was not mined, but panned from rivers in the local area. Villagers in parts of Georgia are still reported to pan the rivers for gold using sheep fleeces — another potential link to the tale of the Golden Fleece. Gold obtained from rivers tends to contain elevated levels of metals such as osmium, tin and palladium, so the archaeologists are now assessing the osmium levels in gold taken from the river to see whether the gold in the artefacts was panned or mined.

Fingerprinting gold is one of the most difficult tasks in archaeometry, says Maria-Filomena Guerra, a physicist at the Centre for Research and Restoration of the Museums of France in Paris, who is separately trying to follow the gold trail. Guerra's team has been analysing gold artefacts from a Mycenaean tomb at Kazanaki near Volos, the modern Greek town close to the area where legend says Jason came from. The tomb dates from 1400 BC and Guerra has so far analysed a gold-beaded necklace and gold leaf interred there, using a technique called PIXE XRF (particle-induced X-ray emission X-ray fluorescence).

Guerra found levels of tin, platinum and palladium in the tomb samples that indicate the gold came from a river source and was almost certainly panned for, she says. She now plans further analyses to prove her hunch that the gold was panned in Georgia. ■

Emiliano Feresin

for many years. He became increasingly unhappy after Springer acquired the journal through its merger with Kluwer Academic Publishers in 2004. Bak says that the editorial board agreed in 2004 to

"There is a lot of concern in the mathematics community about the cost of journals"

consider moving to a different publisher.

Byrne admits there were some teething problems, but says these were being or had been dealt with. She adds that Bak failed to send sufficient manuscripts to fill the journal and that those he did send

arrived erratically, creating difficulties with the workflow. Bak confirms that after April 2006 he stopped sending Springer the manuscripts passed on to him by *K-Theory*'s other editors.

Friedlander is uncomfortable that papers were held up: "Our responsibility is to review mathematics that is submitted to us and disseminate it." Lück calls it "a scandal".

Another former editor says they understood the backlog at Springer to have been so bad that the withholding of the papers would have made little difference to when they appeared. Byrne says: "This is definitely a misconception."

Lück and Ranicki have asked authors to contact them if they have papers accepted

by *K-Theory* that they want to be published in the journal. Meanwhile, Bak has offered authors publication in the new *Journal of K-Theory*.

K-theory, which is concerned with the algebraic properties of space, has had some high-profile moments. One major contributor to the field, Hyman Bass of the University of Michigan, received a US National Medal of Science on 27 July. But K-theory remains a relatively small discipline, which does not need more than one dedicated journal.

For this reason, Ranicki and Lück said on 17 August that they will not try to continue *K-Theory*, but deal only with the papers that have already been accepted. ■

Jenny Hogan

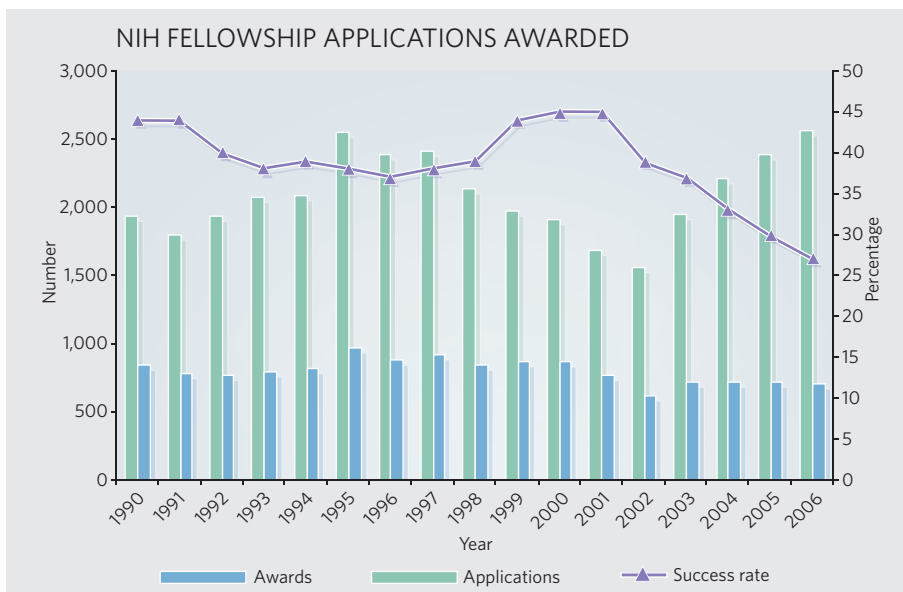
More biologists but tenure stays static

It's official: biology postgraduates in the United States face greater competition for tenure than ever before. A wealth of data released this month will reopen discussions about employment and training in the US biomedical system.

The data, compiled by the Federation of American Societies for Experimental Biology (FASEB) (opa.faseb.org/pages/PolicyIssues/training_datappt.htm), are from many sources, including the US National Science Foundation (NSF), the Council of Graduate Schools and the National Institutes of Health (NIH). And one message is clear: increasing numbers of bright young students are eager for a career in biology and biomedicine, but fewer than before will gain the coveted tenured academic positions.

NSF data show that the number of students in US graduate programmes in the biological sciences has increased steadily since 1966. In 2005, around 7,000 graduates earned a doctorate. But the number of biomedical PhDs with academic tenure has remained steady since 1981, at just over 20,000. During that period the percentage of US biomedical PhDs with tenure or tenure-track jobs dropped from nearly 45% to just below 30%.

So where do all the graduate students end up?



The percentage of biomedical PhDs going into industry has tripled, from 10% to 30%, since the 1970s, the NSF reports. But those who stay on the academic track face a more arduous slog than their mentors. Although numbers of applicants

for postdoctoral fellowships awarded by the NIH increased between 2002 and 2006, the percentage who were successful dropped sharply (see graphic above). And the average age of scientists earning their first R01 grant — the NIH's broad-

Is baby DVD research Mickey Mouse science?

The Walt Disney Company is going toe-to-toe with the University of Washington in Seattle after a study by the university's researchers suggested that exposure to DVDs and videos for babies, including Disney's Baby Einstein, could be associated with poorer language development.

One of the team, professor of paediatrics Dimitri Christakis, was widely quoted as saying "I would rather babies watch *American Idol* than these videos."

Robert Iger, Disney's chief executive, says the study's "methodology is doubtful, its data seem anomalous and the inferences it posits unreliable". In a letter to the university, he accused it of issuing a "deliberately misleading, irresponsible and derogatory" press release. "Whether your university is comfortable associating its name

with analysis of this quality is, of course, your decision," he wrote. "And I would not be reaching out to you if all that was at stake was a poorly done academic study."

What is at stake is a million-dollar industry in such products for babies. A full set of Baby Einstein DVDs costs US\$369.99. Baby Einstein packaging says it "is not designed to make babies smarter", but detractors claim such products are marketed as educational.

"Disney is expecting the brand to bring in \$1 billion by 2010," says psychologist Susan Linn of Harvard Medical School in Boston, Massachusetts, and a founder of the Boston-based Campaign for a Commercial-Free Childhood, a group opposed to marketing to children. In May 2006 the campaign asked the US Federal Trade Commission (FTC) to investigate

Baby Einstein and Brainy Baby of Alpharetta, Georgia, another leading manufacturer of baby videos, for "engaging in deceptive acts and practices". They were backed by the American Academy of Pediatrics and the American Academy of Child and Adolescent Psychiatry. The complaint is still being considered.

Linn sees the entire 'baby industry' becoming more litigious. "We can expect more of this kind of corporate intimidation," she warns. "Disney is on the defensive and they're going to come out swinging."

The study, published online earlier this month (F. Zimmerman *et al. J. Pediatr.* doi:10.1016/j.jpeds.2007.04.071; 2007), found that babies aged 8 to 16 months who watched such videos scored lower than other babies on the Communicative Development

Inventory (CDI), a standard tool used to gauge language development in infants. Babies that watched an hour a day scored 17 points lower on the CDI scale — corresponding to knowing seven fewer words than a typical baby in the study who did not watch the videos, the researchers say.

Lead author Frederick Zimmerman suggests several explanations for their findings, including the fact that parents worried about their child's language development might turn to the videos. But "it is possible that heavy viewing of baby DVDs/videos has a deleterious effect on early language development," he says.

The study was press-released by his university under the headline "Baby DVDs, videos may hinder, not help, infants' language development". University president

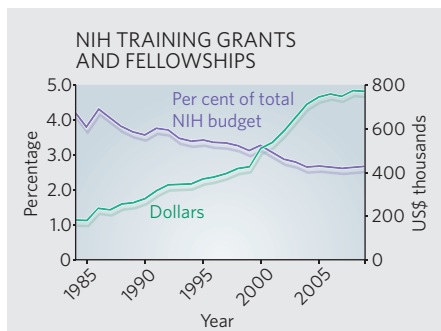


MICE CAN SMELL GREENHOUSE GAS

Rising CO₂ makes a stink for mice.

www.nature.com/news

PUNCHSTOCK



and-butter grant to an independent researcher — has risen from 34 in 1970 to 42 now.

The percentage of PhDs still in a postdoctoral fellowship three or four years after their doctorate has declined since 1997, from 45% to 30%, although the total number of postdocs grew from about 25,000 to 33,000 in the same period.

What does this mean for biology and biomedicine as a career option? It's more than an abstract question to Howard Garrison, director of FASEB's office of public affairs, who has a college-bound daughter. "She's thinking of biological sciences, so I tell her don't give up, but make sure you're realistic about your future," Garrison says. "People are drawn to the biological sciences because it's an exciting field and an exciting time, but people have to have a broad and flexible approach to their careers."

A huge question is why the doubling of the

NIH budget from 1998 to 2003 seems not to have helped young scientists. Most of the money went into infrastructure rather than tenure-track jobs (see *Nature* 443, 894; 2006). The NIH dollars devoted to training fellowships did grow, but not as fast as the rest of its budget, and when the NIH budget stopped growing, so did the dollars (see graphic on left).

Most NIH-supported postdoc appointments are now financed by research grants, not training grants. "If you're being supported to do research on research grants, but you're still a postdoc, there's a tension, because you're not being funded to be trained," says Jodi Lubetsky of the Association of American Medical Colleges.

Norka Ruiz Bravo, deputy director for extramural research at the NIH, says the agency welcomes the FASEB report: "FASEB has performed a very useful and timely service for the biomedical research community in highlighting this important issue. It is a matter of great interest and concern for NIH." The NIH recently instituted the Pathways to Independence awards, which help postdocs set up their own labs.

A posting to an online careers discussion group puts the matter bluntly: "If you aren't thinking about 'alternative careers' before ever setting foot in graduate school, then you're being foolish."

Erika Check

See Editorial, page 839.

Let down by the statistics

Most claims that men and women are affected differently by disease-associated gene variations are poorly founded. A team of researchers has found that the data supporting such claims are often poorly analysed statistically or come from studies that were not adequately designed to show these links.

"The abysmal standard of statistical analysis in much of genetic epidemiology is little short of scandalous," says David Balding, professor of statistical genetics at Imperial College London, UK, who was not involved in the study. "This paper reveals an entire industry of prominently reported results that are largely unjustified and probably mostly false."

John Ioannidis and his colleagues at the University of Ioannina School of Medicine in Greece evaluated 432 claims in 77 research papers (N. Patsopoulos *et al.* *J. Am. Med. Assoc.* 298, 880–893; 2007). The team applied a set of criteria to determine whether the papers' authors had performed the correct analysis, such as comparing like with like, and had taken steps to show that the association was not due to chance. Worryingly, only 12.7% of claims satisfied these criteria. "There is quite a gap between what should have been done and what the journals and reviewers should have asked for, compared with what the authors did," says Ioannidis.

Many studies were not designed to test for a link between sex and gene variants, with researchers trying to extract associations from their data after the fact. Sample sizes were at least ten times smaller than they needed to be to yield statistically robust results, Ioannidis adds.

"This paper tells us that we don't have a clue whether gender is a real biomarker for any of the clinical areas assessed," says Howard McLeod, director of the UNC Institute for Pharmacogenomics and Individualized Therapy in Chapel Hill, North Carolina. "Gender, as well as age and race, are crude ways of understanding the complex factors regulating clinical effect," he adds.

Claire Ainsworth



What babies watch: it may not make a difference.

Mark Emmert has refused to retract the press release and says he stands behind the research. "The findings were



considered significant enough to be reported in a major journal, and as a public institution we feel duty bound to make the public

aware of these findings," he says.

Deborah Linebarger, an expert in child development and television at the University of Pennsylvania in Philadelphia, was asked by Disney to defend Baby Einstein to the FTC. Although believing such products can be useful, she declined. "I have concerns that anything called Baby Einstein, Genius, etcetera, is exploitive

of a vulnerable population," she says.

Despite having "some methodological issues" with the paper, she says: "There are some valid conclusions in it that warrant additional research. I'm cautious, but it makes sense."

Daniel Cressey

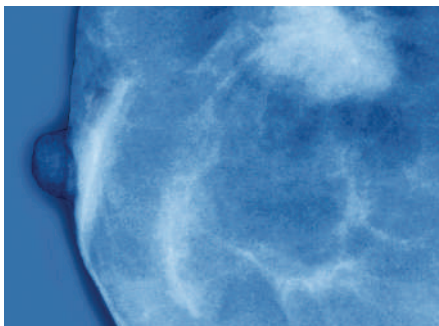
AMERICAN IDOL/19 TELEVISION/FOX/FREEMANTLE MEDIA/KOBAL

Hope for axed cancer-prevention trial

Researchers behind a major trial on breast-cancer prevention whose funding was cancelled in June say that Elias Zerhouni, director of the US National Institutes of Health (NIH) has indicated that he may be willing to meet them.

The trial was to have received US\$55 million from the National Cancer Institute (NCI). It would have involved almost 13,000 post-menopausal women at a high risk of developing breast cancer, and aimed to investigate the efficacy of two cancer drugs — letrozole and raloxifene — in preventing the disease. It was initially approved by NCI staff, external reviewers and, in January, by an executive committee of senior NCI leaders. But the next day, NCI director John Niederhuber put the funding on hold. He convened an ad hoc committee to study the costly trial further.

Niederhuber later cancelled the funding, partly over fears of giving healthy women pills with known side effects and the public's reluctance to take preventative medicines. Previous



Drugs to prevent breast cancer: worth the cost?

studies, he wrote, “have failed to change the practice of breast cancer prevention among women and their [doctors], perhaps because of the side effects of these drugs”.

The scientists leading the trial at the National Surgical Adjuvant Breast and Bowel Project, based in Pittsburgh, Pennsylvania, protested vehemently, and accused Niederhuber of “acting

unilaterally”. On 20 July, they hand-delivered a 13-page appeal to Zerhouni, asking him to intervene. Last week they said, “we’ve had some very positive interactions” with Zerhouni’s office, and they are hopeful of meeting him this month.

The appeal has drawn media attention in the United States, and in a recent editorial *The Lancet* labelled Niederhuber’s actions “troubling” and called for an investigation (*Lancet* 369, 2134; 2007).

But Bruce Chabner of Massachusetts General Hospital, who chaired Niederhuber’s ad hoc committee, says that the NIH’s frozen budget has forced Niederhuber to make tough fiscal choices. He adds that the panel was strongly influenced by the fact that letrozole, made by Novartis, goes off-patent in 2011. “Was it worth spending millions on a trial that might never result in a drug being used for this indication?” he asks. “We thought it likely it wasn’t.” ■

Meredith Wadman

ZEPHYR/SPL

Field trials for GM food get green light in India

India's first genetically modified (GM) food crop is a step closer to reaching the dining table. The government has approved field trials for a strain of brinjal (aubergine) carrying a *Bt* (*Bacillus thuringiensis*) gene.

Jalna-based Mahyco, the Indian venture of US seed giant Monsanto, claims its insect-resistant variety gives better yields with less pesticide use. To avoid possible cross-contamination with farmers' crops, the trials will be carried out in government farms. But critics already campaigning against *Bt* cotton — currently the only GM crop grown in India — say the brinjal trials are illegal.

Full biosafety data on the brinjal tests have not yet been generated, says Kavitha Kuruganti of the Centre for Sustainable Agriculture, based in Hyderabad: "The trials would open doors for entry of GM tomato, okra, mustard, corn, rice and five other food crops in the pipeline." The government says that a committee of experts reviewed all available biosafety data and concluded they were in order.

Gates Foundation helps Iraqis escape the conflict

The Bill & Melinda Gates Foundation is getting into the business of helping persecuted Iraqi academics continue their research abroad. A \$5-million grant will help support the Scholar Rescue Fund run by the Institute for International Education (IIE), which aims to raise \$15 million for Iraq.

The Gates Foundation is best known for its \$33-billion bankroll and for tackling health problems such as malaria and HIV. The new grant is a one-time award, not a shift in focus, says a spokesperson for the foundation.

As Iraqi professors come under increasing threat of violence, many have sought help to temporarily flee the country, says IIE president Allan Goodman. So far, 41 grants have been awarded to Iraqis — 27 to



Car-bomb wreckage near Baghdad University highlights the perils for some Iraqi academics.

Asthmatics win payment in diesel-fumes lawsuit

Asthma patients in Tokyo last week welcomed a cash settlement from car manufacturers and the Japanese government. The one-time payment resolved a decade-long legal battle in which the asthmatics blamed diesel car fumes for their illness.

The automakers, including Toyota, Honda and Nissan, will provide ¥1.2 billion (US\$10.5 million) to the plaintiffs, and a further ¥3.3 billion to support a five-year health plan for the patients. The central government and Tokyo metropolitan government will each contribute ¥6 billion to the medical programme.



scientists — to pay for relocation, housing and research. "This is about the rescue of science and learning," says Goodman.

WWF sues over threat to endangered whales

Last month, in the Sea of Okhotsk, Russia, a conglomerate of energy firms lowered a 28,000-tonne platform for drilling for oil on to a base on the sea floor. They did it very carefully — in part to keep down the noise, as the last feeding grounds of the western Pacific population of grey whales (*Eschrichtius robustus*) lie some 7 kilometres away.

But environmentalists and some scientists say construction of the platform was not quiet enough. So the environmental group WWF and activist group the Corner House have taken legal action against a UK government department. The WWF claims that the department conditionally promised financial support for part of the Sakhalin-II project without approving its environmental impact.

A department spokesperson calls the case "flimsy" and points out that the environmental review was in fact one of the conditions that would have to be met for approval. It is estimated that the Okhotsk region harbours the equivalent of 45 billion barrels of oil.

Anticlotting agent gets gene-specific label change

The US Food and Drug Administration (FDA) is changing the label on a widely used drug to include information about how two genes may affect dosage requirements.

The FDA said on 16 August that the label on warfarin, a popular blood-thinner with the trade name Coumadin, is to be revised. The new version will state that a lower

starting dose "should be considered for patients with certain genetic variations".

Warfarin is widely used in patients with artificial heart valves and those at risk of blood clots. But it can cause massive and sometimes fatal bleeding. The two gene variants flagged by the FDA encode enzymes involved in warfarin's breakdown and its function as an anticlotting agent.

Although the FDA has demanded similar labelling on four other drugs, warfarin is the first major drug to be labelled in this way. "This means that personalized medicine is no longer an abstract concept but has moved into the mainstream," says Larry Lesko, director of the FDA's Office of Clinical Pharmacology.

India to expand research institutes and universities

The Indian Planning Commission has rolled out a seven-year plan to set up eight more Institutes of Technology — the elite science and technology universities. Also in the US\$32-billion plan are 7 new national business schools and 20 Institutes of Information Technology.

The push follows a call by Prime Minister Manmohan Singh to put Indian science on the global map by upgrading its research facilities. There are also plans for 30 more 'central universities' — one in each of the 16 states with no such institution at present, and 14 in states that offer free land.

Notice to authors and readers

Nature and the *Nature* journals have refined their policy on competing financial interests. *Nature* requires that all authors declare that they do or do not have competing financial interests relevant to their submission. If an article in *Nature* does not contain such a statement, readers can assume that authors declared to the editors that they have no competing financial interests. See www.nature.com/authors/editorial_policies/competing.html for full details of the policy.

BUSINESS

A change of gear at Siemens

Europe's largest technology company is switching its priorities, even as it faces a growing bribery scandal. **Quirin Schiermeier** reports.

Almost half-a-million people work for Siemens, the Munich-based electrical-engineering company with interests in everything from bullet trains to medical diagnostics. So when a bribery scandal hit the company late last year — forcing out several of its top managers — the stakes for German science and technology were high.

Siemens has emerged over the past half-century as an icon of German engineering. But its new chief executive, Peter Löscher, is the first in the company's 160-year history to be brought in from the outside. Markets have not responded particularly well to his July appointment (see graph), and new disclosures about the extent of the bribery scandal are weighing heavily on the company.

Löscher, a former manager with the pharmaceutical company Merck, is an unknown quantity at Siemens. But the 49-year-old Austrian, who is described as determined and down-to-earth, hasn't wasted much time on making an impression: just last month, the company struck two major deals that, analysts say, mark a considerable change in the direction of the company. It is getting out of a stable, low-margin business — auto parts — and deeper into a dynamic, high-margin one — medical diagnostics.

Siemens goal is to be “the world's first full-service medical-diagnostics company”, Löscher said last month. Developing easy-to-handle ‘integrated’ diagnostics systems is becoming a cornerstone of its research strategy, adds Reinhold Achatz, the electrical engineer who directs the ten central research laboratories that the company operates around the globe.

Achatz says that the company's drive into medical diagnostics will aim to optimize hospital workflow by integrating the results of procedures such as computerized tomography with those of laboratory tests. “Advanced diagnostics produce an enormous amount of medical information,” he says. “Translating this information into early diagnosis of diseases requires strong knowledge management and very efficient processing technologies.”

Siemens says it spent some €5.7 billion (US\$7.7 billion) on R&D in 2006 — more, it says, than IBM, Matsushita or, indeed, any rival company. Its annual sales of €87 billion

rank it second in the electrical and electronics engineering sector, behind General Electric (€129 billion).

The research carried out by the 1,900 staff in Achatz's labs lies at the more ‘basic’ end of the spectrum, and accounts for just 5% of the Siemens' overall spending on R&D. The company employs a total of almost 50,000 research and development staff, including some 30,000 computer programmers, based in 30 countries.

Collaboration with universities is an important part of the company's research strategy, Achatz says. Each year, more than 1,000 projects are launched in conjunction with academic researchers all over the world. Siemens initially provides money, infrastructure and a business plan if there is possible commercial potential. After a year or so, the firm decides whether or not to go forward with the project. “That way we make sure that no ideas get lost,” says Achatz.

Home-grown success

Germany remains the bedrock of innovation in the company, with almost half of the research staff based there. Many pioneering inventions, from electric locomotives in 1879 to whole-body magnetic resonance imaging in 2004, have been conceived in Siemens' German labs.

But with Asian markets becoming increasingly important, Siemens' research footprint is going global. “Some 300 million people in China and India may already buy and use our products, but 2 billion others don't,” says Achatz. The company has hired about 300 scientists and engineers over the past five years at labora-

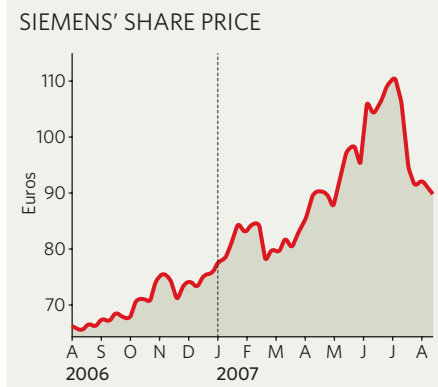
tories in Beijing, Shanghai and Bangalore.

The rise of Asian technology corporations poses a strategic threat to Siemens, which it hopes to counter with its extensive know-how based on operations in 190 countries. “We have a competitive advantage in that knowledge from different cultures leaves its mark on our products,” says Achatz.

He adds that Siemens' researchers haven't felt much impact from the bribery allegations that have shaken up other parts of the company. “We are very aware that trustful partnerships are of the essence when contracting out research,” Achatz says. “All of our deals need to be legally impeccable.”

The problem first emerged last November, when police searched Siemens offices in Germany and Austria for evidence that corporation officials had bribed customers and trade-union negotiators. Sham businesses in several European countries came to light, and around €420 million of suspicious payments are currently being reviewed by lawyers and public prosecutors. But according to a report published last week by the newspaper *Süddeutsche Zeitung*, internal reviewers have found evidence that dubious transactions amount to more than €1 billion — and that several divisions of the company are affected.

Two board members were temporarily imprisoned last winter. And in April, when the US Securities and Exchange Commission announced its own investigation, chief executive





Clinical view: medical diagnostics is a critical area of expansion for Siemens.

Klaus Kleinfeld and chairman Heinrich von Pierer were forced to step down.

Kleinfeld — who was chosen last week to run the aluminium company Alcoa — had already sought to reduce the range of company activities, giving up, for example, on the mobile-phone business. Siemens shares had doubled in value during his watch.

Löscher must now complete the restructuring process while restoring the company's reputation. He agreed to buy Dade Behring, an Illinois-based medical-diagnostics supplier, for \$7 billion on 26 July — and to sell VDO, Siemens' auto-parts unit, for \$7.6 billion on the same day. The markets didn't like the deals much, believing that Siemens could have done better on the pricing of both.

Siemens had already purchased German company Bayer's medical-diagnostics unit for almost \$6 billion in 2006, as well as two US companies in the same sector. The acquisition of Dade Behring makes Siemens the second-largest player, after Roche Diagnostics of Switzerland, in a global diagnostics market worth some \$60 billion.

Although the market has been tough on Siemens over the past month, most analysts accept that acquisitions make sense in the long term, and will improve the firm's strategic position in medical technology. That, they believe, will help Siemens to maintain its place as a lynchpin of the German economy and one of the world's most successful technology companies. ■

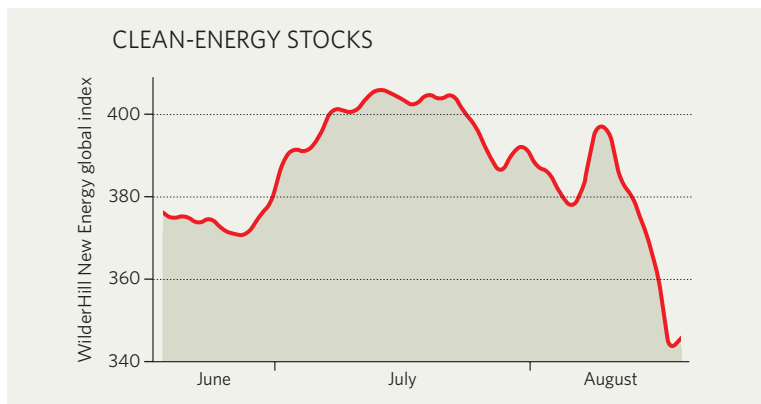
IN BRIEF

BIOTECHNOLOGY BLOW Amgen says it will cut between 2,200 and 2,600 jobs in response to falling sales of its anaemia drug, Aranesp (see *Nature* 447, 899; 2007). These will be the first-ever layoffs at the Thousand Acre, California-based company, whose phenomenal rise has — together with that of Genentech in San Francisco — provided the dominant corporate model for the biotechnology industry worldwide. But Amgen has come under pressure from Wall Street to cut costs: announcing a restructuring on 15 August, the company said that the proportion of revenue spent on R&D would drop from last year's 23% to about 20%.

WIND-ENERGY BOOM The United States is set to install more than 3,000 MW of wind-power capacity this year, up from 2,450 MW last year, according to figures compiled by the American Wind Energy Association. Total installed capacity now stands at about 12,600 MW: just over 1% of total US electrical generating capacity. Analysts say that the wind-energy market is maturing, with lenders and investors increasingly happy to support new projects.

PAINKILLER PULLED A Novartis painkiller that is a rare surviving member of the COX2 inhibitor class of drugs has been withdrawn from the market in Australia, where an estimated 60,000 people were taking it. Australian regulators cancelled the registration of the drug, Prexige (lumiracoxib), on 11 August, saying that they had received eight reports of serious liver toxicity, including two deaths and two people who required liver transplants. The Swiss company sells the drug in over 50 countries including the United Kingdom, and is applying for marketing approval in the United States (see *Nature* 448, 400–401; 2007.)

MARKET WATCH



The global plunge in stock markets over the past three weeks has taken the wind out of the sails of clean-energy stocks, ending a powerful run that saw their value grow by some 40% in the year to mid-July.

The WilderHill New Energy Global Innovation Index (NEX), which tracks a number of companies around the world with strong interests in renewable energy, plateaued in July and then tumbled this month, along with the general stock markets. Some observers say that, for clean-energy stocks, this was an overdue correction.

The index has been hit hard by the market turmoil, says Michael Liebreich, chief executive of London-based New Energy Finance, which co-compiles the index with WilderShares of California. "It's taking a little bit of a beating," he says, "but it's still ahead on the year." The

index started 2007 at 289 and hit 400 before retreating this month (see above).

The early summer had been particularly strong for US clean-energy companies, which had (and indeed still have) high hopes that the energy bill winding its way through Congress will provide fresh incentives for investment in wind and solar power. Rob Wilder, president of WilderShares, says that just before this month's general market slump, investors were betting that the bill would pass — and boost demand for renewable energy sources in the United States.

"We expect to see further turbulence and volatility," says Liebreich, who has been warning for some months that the index was due to dip (see *Nature* 447, 17; 2007). "But the fundamentals of the industry are very strong. In a way I actually welcome the correction." ■

Colin Macilwain



Determined stance: Robert Place (left) and Long-Cheng Li (right) vow to continue their controversial work on RNA activation.

Hitting the on switch

Researchers in San Francisco have findings that suggest a whole new side to RNA interference. **Erika Check** reports on their attempts to make a revolutionary field more revolutionary still.

"That looks perfect," murmurs Robert Place as he watches a smooth line trace across his monitor. "It never works this well."

Place, a molecular biologist at the Veterans Affairs Medical Center and University of California, San Francisco, is using a spectrophotometer to measure the purity of a series of samples. Every sample contains a tiny drop of micro RNA (miRNA), a type of genetic regulator that dampens gene expression — or so the story goes. The experiment Place is absorbed in is the last he must complete before he submits a publication that could upend that story. He and his colleague, Long-Cheng Li of the University of California, San Francisco, think they have found some miRNAs that boost, rather than silence, gene expression in cells.

Their work could shake the foundations of one of the hottest topics in biology — RNA interference — which studies how short pieces of RNA regulate the expression of genes. Place knows that his experiment will draw intense

scrutiny from other researchers, and therefore it has to go perfectly. And as far as he can see, it has. His spectrophotometer displays a series of flawless curves, free of impurities; it looks as if he and his colleagues are finally seeing the pay-off after a three-year saga of frustration and exhilaration.

Since about the turn of this century, scientists have realized that 50 years of focus on DNA had blinded them to the wide range of biological roles held by its chemical cousin, RNA. The old view was that DNA contained life's instructions, proteins carried them out, and RNA served as little more than a go-between. It's now become clear that RNA has vast potential for controlling how cells interpret the instructions embedded in the genome.

The RNA revolution began in 1998 with the discovery honoured by last year's Nobel Prize

in Physiology or Medicine that small strands of RNA stuck together in pairs, like the strands that make up DNA, could turn off specific genes in roundworms¹. RNA interference was born.

In 2001, scientists discovered that the process works in mammals, too². They found that interference could be triggered by 'short interfering RNAs' (siRNAs) or 'short hairpin RNAs' (shRNAs).

Both work by using a pair of scissors, actually a complex of proteins known as RISC, to cut apart the longer messenger RNAs (mRNAs) that take information from the genes to the cell's protein-making machinery. The small RNAs target mRNAs because their sequences match, and by destroying the messenger, RNA

interference stops genes from making proteins. Both siRNAs and shRNAs are being tested in clinical trials against conditions in which an overactive gene needs to be shut down³.

"It was so easy to observe activation that I didn't understand how people could have ignored it."

— Long-Cheng Li

Scientists then discovered miRNAs, which also trigger the interference pathway, but unlike siRNAs and shRNAs are naturally encoded by cells' DNA. So far, more than 500 miRNAs have been identified in the human genome⁴. They can act like siRNAs, using the same protein complexes that slice through messenger RNA. But for the most part, miRNAs are content to muzzle the message or mark it for degradation, rather than chop it up (see graphic).

All these techniques were still quite new back in 2004, when Li began to investigate RNA at the San Francisco Veterans Affairs Medical Center in the laboratory of Rajvir Dahiya, a urologist. Li, also a urologist, was studying epigenetics — stable modifications to the genome that change how it is read without altering its sequence.

He was particularly interested in DNA methylation, the addition of chemical tags called methyl groups to regions of DNA. Methyl tags

often silence nearby genes, which can be disastrous for genes that usually suppress tumours. So Li was trying to find ways to reverse this process. He decided to try an RNA-based technique to control methylation that hadn't yet been used in animals, as far as he knew, although it had been demonstrated in plants a decade earlier⁵.

Unexpected boost

Li purchased two pieces of double-stranded RNA that were complementary to a DNA control sequence — known as a promoter — upstream of the gene that encodes the tumour-suppressor protein E-cadherin. Then, his colleague Hong Zhao added the double-stranded RNA to two lines of prostate cancer cells and measured how it affected E-cadherin expression. Three days later, Zhao noticed something shocking: the experiment had boosted levels of the E-cadherin protein by 4- to 14-fold. "I couldn't believe it," Li says.

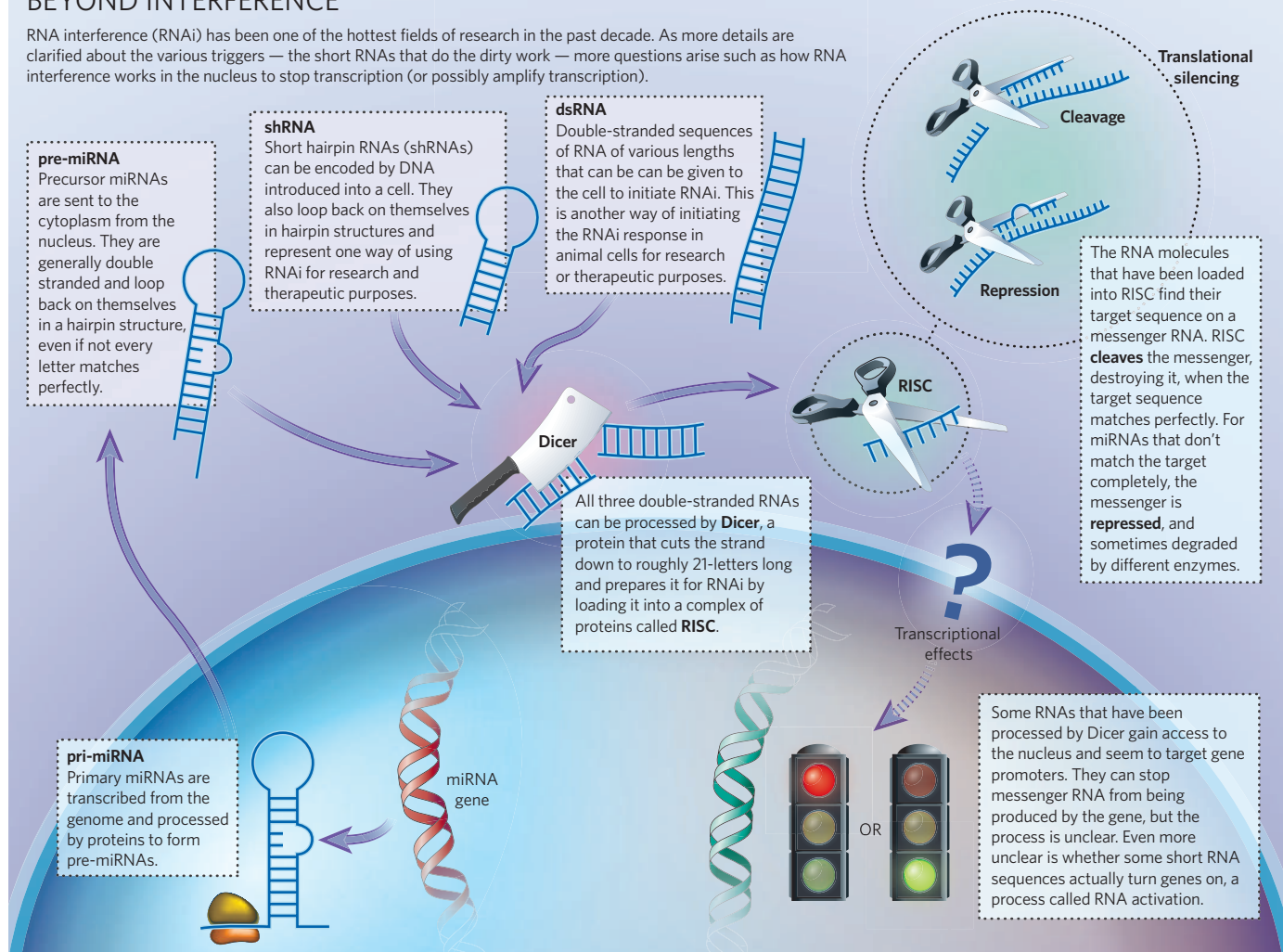
The work defied everything that scientists had ever reported about small, double-stranded RNAs. They were only supposed to dampen gene expression. Li seemed to have stumbled on a phenomenon that could rewrite the textbook understanding of RNA interference and might offer new therapeutic potential.

Li re-ran the experiment many times, each time with the same result. He found that two other cancer-related genes, *VEGF* and *p21*, could also be activated by double-stranded RNAs. The evidence seemed solid. At least in the prostate cancer cells he was working with, short, double-stranded RNAs activated genes. "It was so easy to observe — I didn't understand how people could have ignored it for such a long time," Li says.

But apparently they had. So Li knew he was going to have to put together a watertight case to convince other scientists. He started to realize how difficult that would be when he first submitted his work for publication to *Science* in August 2004. It was promptly

BEYOND INTERFERENCE

RNA interference (RNAi) has been one of the hottest fields of research in the past decade. As more details are clarified about the various triggers — the short RNAs that do the dirty work — more questions arise such as how RNA interference works in the nucleus to stop transcription (or possibly amplify transcription).



W. FERNANDES

rejected. He then submitted his paper to *Nature* that December; when it was rejected, he resubmitted it with new data in April 2005. He presented his findings at the annual conference of the American Association for Cancer Research in Anaheim, California, in May 2005, and the reception wasn't warm. "I got a lot of sceptical questions," Li says. Then, after an extensive delay, *Nature* again rejected Li's paper in December, 2005. Without evidence for a mechanism, he says he was told the results weren't convincing enough.

Unsure how to proceed, Li got some unexpected help in the form of Place, who joined the lab in October 2005. Fresh from earning his PhD, Place was excited by the novelty of Li's findings. But he could tell that the lab had a lot of nitty-gritty molecular biology work to do. Place helped plan experiments to help convince the sceptics. For instance, he helped Li and his technician Deepa Pookot to perfect assays, called immunoblots, that detect proteins so that they could measure how RNA activation boosts levels of E-cadherin and other proteins.

Filling in the details

Place and Li also started to think about the mechanism behind RNA-directed gene activation, a major stumbling block for the editors and reviewers who had seen the paper. If it worked in the same way as gene silencing, the sequence of the trigger RNA should matter, and the pair found that it did. Changing five letters at one end of the 21-letter sequences rendered them inactive.

The pair also found that activation used some of the same key proteins that are involved in silencing, such as Dicer, which cuts up strands of RNA so that they can be used by RISC to target mRNA. They began to experiment with activation, trying to find out what worked best. "We started to modify the RNA duplexes to get optimal activation, tweaking their chemical structures," Place recalls. "There came a point when we were working in synchronization, and we really started to click."

Li resubmitted the work to *Science* with Place's additional molecular biology results, but it was again rejected. The letter he received said that because the work "would represent a substantial paradigm shift", the evidence just wasn't strong enough. Again, editors required demonstration of a mechanism.

So, in June 2006, Li talked to leaders of the RNA-regulation field at a meeting in Cold Spring Harbor Laboratory, New York. Li recalls

"Last time I talked about activation at a meeting, a couple of leaders in the field jumped all over me and said it wasn't true."

— David Corey

us," Place says. "Before that, it seemed like everyone just thought we were crazy."

Back at the Veterans Affairs Medical Center, Place and Li were now working closely together, although Li was preparing to move to his own lab at the University of California, San Francisco. They had already devised a series of experiments that they hoped would bolster their activation hypothesis. Although some changes to the sequence can render RNA useless, others fine-tune the activation. The trigger doesn't

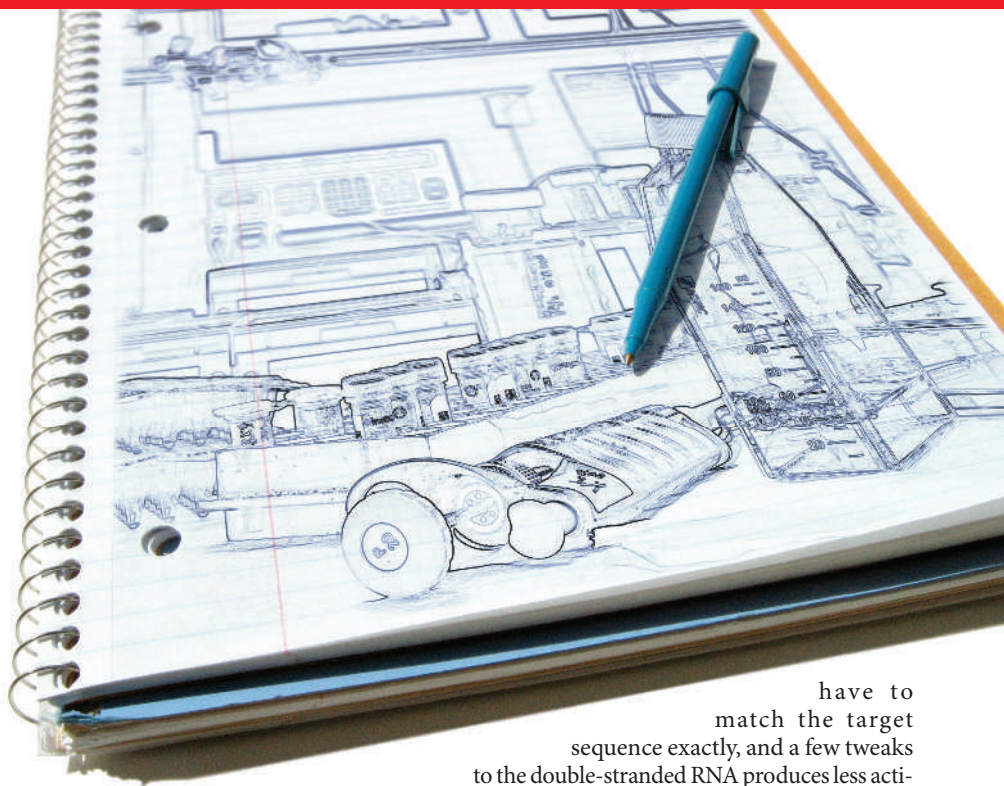
have to match the target sequence exactly, and a few tweaks to the double-stranded RNA produces less activation. And that's exactly how miRNA works. So the pair started hunting for a miRNA that might activate E-cadherin. It wouldn't prove that activation was a natural phenomenon, but it would strengthen the case, because miRNAs are encoded in the genome.

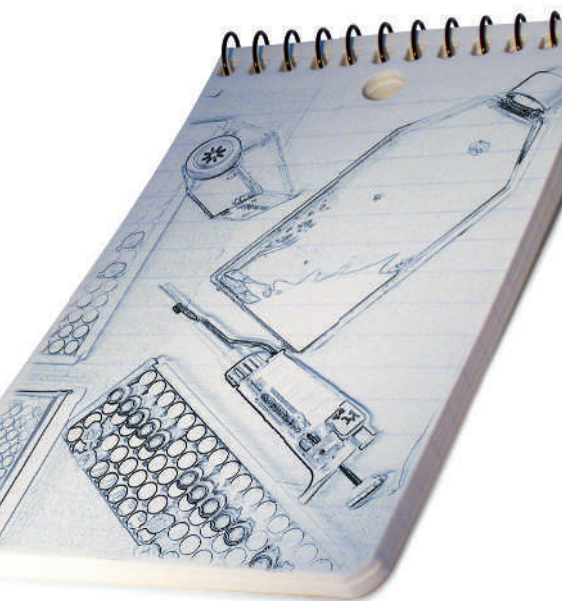
Place and Li used a bioinformatics tool to search for miRNA in the human genome that had sequences roughly complementary to the E-cadherin promoter sequence. They found a handful of candidates and transferred them into the prostate cancer cell lines. The experiment was so simple they half-expected it wouldn't work — but it did. The pair found one miRNA, miR-373, that boosted E-cadherin expression. "When that worked, we were totally pumped, because it was a potential example of natural function," Place says.

On the right track

But it still wasn't proof. So Place designed an miRNA precursor that, according to the model of miRNA biogenesis, should work as well as the miRNA itself. It did. Place also found that if he knocked out Dicer, then miR-373 stopped working, and that Dicer could also activate another protein with a promoter sequence similar to E-cadherin's. All these experiments supported the idea that miRNA could use the interference pathway to activate genes. It was a strong hint that Li and Place were on the right track. On 6 August, they submitted the work to *Nature Chemical Biology*.

The question now is how this work will be received. Many are still sceptical about activation, as David Corey can attest. "Last time I talked about this at a meeting, a couple of leaders in the field jumped all over me and told me it wasn't true," Corey says. "We've received grant and manuscript reviews that seem to express irritation more than anything else. I





C. PICKENS

just wish they would look closely at our data.”

Indeed, it sometimes seems as though the idea of activation is struggling because it contradicts the interference dogma. The RNA interference field is quite young, but already seems to have acquired a certain amount of inflexibility. On scanning through archives of biology message forums, Place has found that other scientists — often graduate students — have also seen evidence of RNA activation. But they have been encouraged to discount it.

Phillip Sharp, a Nobel-prize-winning biologist whose lab at the Massachusetts Institute of Technology in Cambridge pioneered much of the work on RNA interference, admits that RNA-mediated activation might be possible, but says that “the results I have seen do not prove this”. Echoing the demands of manuscript editors, he says, “I think further papers on the subject must address the mechanism if they are to be published in a high-profile journal.”

A new pathway

As the field has moved apace, such requests are warranted, and providing proof of a mechanism is Li and Place’s main challenge. Some have suggested that activation is simply inhibition in disguise. It could be the accidental result of silencing an upstream repressor or of blocking another silencing RNA. Although they can’t rule out these possibilities, from their experiments, they say that these mechanisms look unlikely. For instance, they can elicit activation predictably at specific genes by targeting their promoters.

“Further papers on the subject must address the mechanism if they are to be published in a high-profile journal.”
— Phillip Sharp

In this together: Long-Cheng Li (left) is briefly reunited with the team from the Veterans Affairs Medical Center. Left to right, Li, Emily Noonan, Robert Place, Deepa Pookot and Rajvir Dahiya.

There are intriguing differences between the known silencing pathways and the observed activation phenomenon. Silencing is triggered within hours and ceases in about seven days, whereas activation takes days to appear but can last for weeks. The different kinetics suggest that some mystery process is involved, Place says. “People say this could be RNA interference with another name, but it’s so blatantly different.”

Major questions also remain about how RNA regulation might act at gene-promoter regions. In the classic interference pathway, the RISC complex guides siRNAs or miRNAs to a target mRNA in the cell’s cytoplasm. But to regulate a promoter, a small RNA would have to sneak into the cell nucleus, where DNA is transcribed. There is mounting evidence that this happens in silencing pathways, but again the mechanism is unclear.

In 2004, two groups published papers that suggested that siRNAs that target gene-promoter regions can silence genes if they are delivered into cell nuclei. Although one group has since retracted its paper, the other group, who published in *Science*⁸, showed that the inhibition was accompanied by epigenetic marks associated with silencing. Kevin Morris, of the Scripps Research Institute in La Jolla, California, an author of that paper, has continued to study how this occurs, and sympathizes with Li and Place’s position. “I was there in

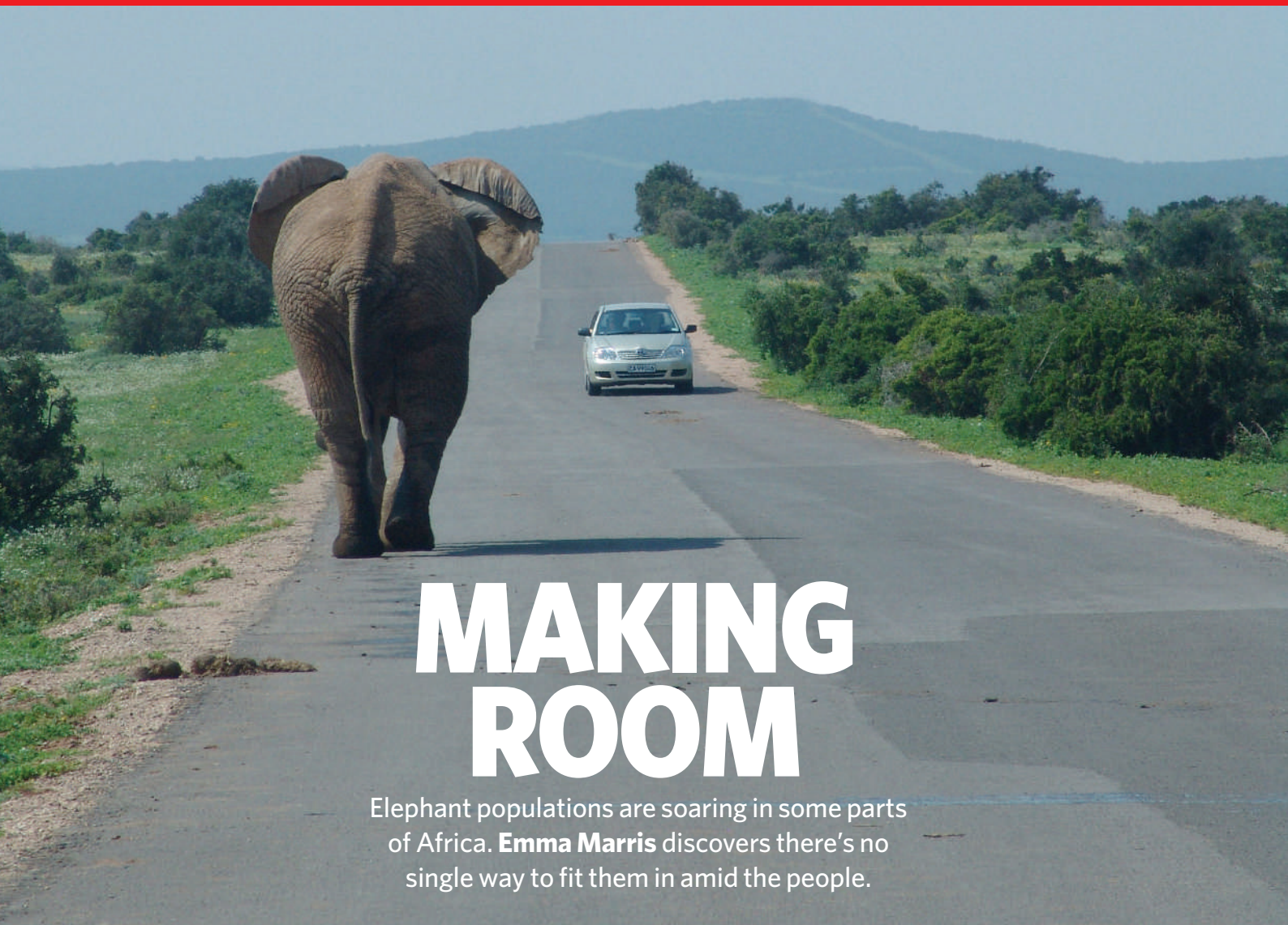
2004, when 50% of the people love your work and the other 50% think you’re full of it,” Morris says. “It’s a frustrating place to be.”

Frustrating, indeed. As the deadline for this news feature approached, Place received word that the group’s manuscript on miRNA activation had been rejected. The reason given: without proof of a mechanism, the evidence isn’t substantial enough. Li is still adamant that the field will come around. “We have no doubt that RNA activation is an endogenous mechanism,” he says. And Place seemed unsurprised. “We knew the mechanism would be the sticking point. That’s the hardest part to prove.” But, he predicted, if the group can just get someone to just take a look at its data, the strength of its evidence will prevail. “If we can get it into review, we’ll be okay,” Place says.

When RNA interference first hit the scientific radar, it was a slow climb from something written off by many as artefact to a revolutionary paradigm. That same uphill battle confronts Place, Li and their collaborators as they try to rewrite, or at least refine, the revolution. Call it what you will — stubbornness, confidence or optimism — this group just isn’t going to give up.

Erika Check writes for Nature from San Francisco.

1. Fire, A. *et al. Nature* **391**, 806–811 (1998).
2. Elbashir, S. M. *et al. Nature* **411**, 494–498 (2001).
3. Check, E. *Nature* **442**, 614–615 (2006).
4. <http://microrna.sanger.ac.uk/sequences>
5. Wassenegger, M. *Cell* **76**, 567–576 (1994).
6. Li, L. C. *Proc. Natl Acad. Sci. USA* **103**, 17337–17342 (2006).
7. Janowski, B. A. *et al. Nature Chem. Biol.* **3**, 166–173 (2007).
8. Morris, K. V. *Science* **305**, 1289–1292 (2004).



MAKING ROOM

Elephant populations are soaring in some parts of Africa. **Emma Marris** discovers there's no single way to fit them in amid the people.

The car creeps towards the left as the driver tries to get as close as possible to the group of elephants foraging at the road's edge. The elephants walk silently, communicating in companionable purrs. The loudest noise is the crackle of the tough foliage they are eating.

"You're crowding Agatha and she's going to have to press into the bush to get by," says Katie Gough from the back seat. Gough, based at Nelson Mandela Metropolitan University (NMMU) in Port Elizabeth, South Africa, has been studying these animals for four years, and can tell most of the elephants apart by the wear and tear on their ears or by idiosyncratic wrinkles. She is right about Agatha, who, slowly moving her enormous body into the thorny shrubs, turns her head and gives the occupants of the car a look that everyone reads — scientific prohibitions on anthropomorphization be damned — as reproachful.

Agatha's home, Addo Elephant National Park in South Africa, has too many elephants. In 1954 there were 22 animals in a park of about 2,300 hectares. They were the remnants of a herd hunted nearly to extinction by one hired hunter in 1919. He was carrying out

the orders of local orange growers who were sick of elephants gorging themselves on their crops. Today, there are around 460 animals in about 26,000 hectares — or roughly double the 1954 density. Throughout its history, densities have waxed and waned; in the main camp of Addo they now stand at about three elephants per square kilometre. One estimate of the maximum number of elephants this area can sustainably support is about 0.5 elephants per square kilometre¹.

"Twenty or thirty years ago, things were black and white," says Graham Kerley, the elephant expert at NMMU who drove too close to Agatha. "You had elephant management and you had *laissez-faire*." Although faced by an excess of elephants, Kerley is clearly still charmed by them, as he is by all the carefully managed animals at Addo — among them the elegant kudu, zebras with tawny rumps, and the immense ostriches sprinting along the road.

These days no one believes in *laissez-faire*. The effects that too many elephants have on their environment are easy to see. Climbing the stairs over a fence to cross from an elephant area into a pachyderm-free zone, the

landscape switches from patchy shrublands to a lilliputian forest containing a wide variety of plants, including spectacular aloes.

Giant appetite

In places where the Addo elephants have already consumed the juicier vegetation, they now feed on bushes such as spekboom, sweet thorn and the bee-sting bush — plants as tough and spiny as their names suggest. In the Addo area, elephants may threaten some 168 species of plants with extinction². These include the geophytes and the aptly named succulents, including the jade plant. And as the miniature forest turns into miniature savannah, other animals will see their habitats change. Animals that live in areas dense with vegetation, such as the cape grysbok and bushpig, will have their habitats fragmented.

As the Addo experience shows, if your goal is to preserve an entire ecosystem and not just the elephants within it, pachyderm numbers have to be controlled. But where park managers, scientists and government experts disagree on the best means of managing elephant populations, the elephants can find themselves left in limbo.

For an outsider, it's a challenge to switch

K. GOUGH

from seeing African elephants as much of the Western world usually does — as majestic, rare animals dying at the hands of villainous ivory poachers — to a more nuanced, more African conception. Depending on who and where you are in Africa, elephants may also be livelihood-destroying crop-raiders, expensive tourist lures, environmental menaces, or dinner.

Different countries, different stories

So it comes as no surprise that Africans can become irritated when told how to manage their elephant populations by interested parties abroad. Well-meaning and sometimes technically expert outsiders are constantly offering opinions on whether to cull animals, whether to try contraception, whether some kinds of ivory should be legally sold, and so on. Often, onlookers will treat the whole continent as a single case, when each country and each area has its own problems to solve.

Most African elephants still live outside protected areas, and so they increasingly come into fraught contact with Africa's expanding population. Botswana has an estimated 150,000 elephants roaming without major barriers between them and pastoralists herding livestock. In Namibia, elephants range free of people in the western deserts, but they butt heads — sometimes literally — with agriculturalists elsewhere in the country. In other countries, elephants have been fenced into reserves or hemmed into protected areas by development as more land is built on.

Across southern Africa as a whole, elephant populations are increasing by about 4% annually, while herds in central and western Africa continue to struggle with poaching and habitat fragmentation⁵. “I could tell you a different story for every African elephant range,” says Holly Dublin, an elephant expert who is chair of the Species Survival Commission at the World Conservation Union.

Nowhere do humans and elephants pack in more tightly than in South Africa. So it's tempting to see the country as an example of how elephant management might look across Africa in the future. But the country is different to others in that there are fewer than half-a-dozen unfenced elephants nationwide. These are the elephants of the Knysna forest, which have held on to their freedom only by virtue of living in an inaccessible valley. All the other elephants are descendants of remnant herds, living behind strong fences built with materials such as railway ties and mineshaft cabling.

“South Africa cannot be considered indicative of the challenges facing elephant conservation

in the rest of Africa,” says Dublin. “It is more like living in America and trying to take care of elephants. For the most part, rural communities are not trying to make their livelihoods right next door.”

South Africa itself has only a small fraction of the continent's elephants. But to judge from media reports, they might as well have almost all of them. Coverage of the culling controversy is largely responsible for this. Until 1994, managers at South African National Parks (SANParks) would shoot elephants in places such as Kruger National Park, far north of Addo on the Mozambique border, whenever their numbers got too high. The slaughtered elephants would be given to local people as a free meal. Other elephants were transported to parks that wanted more elephants or sold to private reserves.

Before the modern practice of culling whole family groups was introduced, culling occasionally resulted in orphans that grew up to behave antisocially. For example, some orphan elephants, which were moved from Kruger to other parks, famously became maladjusted juveniles that went around killing rhinoceroses³.

Getting in a flap

In 1994, culling was stopped because of the objections of local and foreign animal rights groups. Elephant densities have since increased, and in Kruger, where there are now about 0.63 elephants per square kilometre⁴, large bulls flex their muscles by pushing over baobab and marula trees. And after a decade of debate, in 2004, SANParks announced that a whole suite of tools — including culling — should be considered for managing elephant numbers.

This news was not received well, especially by wildlife conservation groups, who claimed that culling was inhumane, even unethical.

Hector Mogame, executive director of Conservation Services at SANParks, Pretoria, doesn't see it this way. “The issue of ethics is about power,” he told the Society for Conservation Biology in Port Elizabeth this July. “The viewpoint of opposition to lethal control is usually favoured by affluent people — people with money.” Citing the furore over SANParks' plan to resume culling, he says the debate is about the people on the ground being trumped by the rich and powerful.

SANParks officials have historically had a relatively free rein to manage the animals within their park boundaries. But after the 2004 outcry, South African government ministers stepped in and asked SANParks to provide scientific justification for its culling plan. In 2005, Marthinus van Schalkwyk, minister of tourism and the environment — and these two are very closely linked in South Africa — called for a scientific round table representing different views on the elephant management debate. Mogame and Kerley were two of those invited.

In January 2006, the Elephant Science Round Table decided that Kruger didn't have enough data at that moment to show culling was necessary, but smaller reserves experiencing “bigger pressures” might need to do something immediate. The group has continued to meet unofficially to advocate for a major research push on elephant management.

SANParks has moved away from suggestions that there is an ideal number of elephants. “The maintenance of biodiversity is best achieved by permitting — or if appropriate actually encouraging

“Opposition to lethal control is usually favoured by affluent people.”
— Hector Mogame



Changing diet: elephants move on to sweet thorn (above) once the jucier aloes (left) have been consumed.



— variation in time and space, rather than attempting to manage for stability,” says Wanda Mkutshulwa, head of communications at SAN-Parks. As of now, the official management strategy of SAN-Parks is “wait and study”.

In the meantime, the elephants are continuing to restructure their habitat. Their movements are dictated by the presence of food and, crucially, water. Around any surface water where elephants drink, you will find a ‘piosphere’, a circle of reduced vegetation cover demonstrating the landscape-changing power of these animals. The shrubs remaining have just a few leaves, and seem as if they are clinging to life. There is no grass, and warthogs are pale from dust bathing. The word ‘desertification’ springs to mind.

Adrian Shrader of NMMU argues that if you pack water-holes and elephants in so densely that the piospheres begin to overlap, landscapes may begin to suffer irreversible losses. But even without such pressures, a high number of elephants will over time reduce plant diversity.

Which path to tread?

For now, park managers still have several other options besides culling that they can actively pursue. But which is best? Most elephant experts have a pet intervention they study or favour. Round-table member Bruce Page, at the University of KwaZulu-Natal, Durban, is keen on contraception, which has promise but is extremely awkward in execution. It requires identifying elephants, and then shooting hormones into them from a truck or helicopter. Field trials at Kruger recorded reasonable success rates.

In Page’s population models, contraception and culling affect population structure in different ways. The proportion of the population under ten years of age increases under culling, but decreases under contraception. When population control ceases, you get a baby boom after culling that you don’t get after the cessation of contraception. Still, Page estimates that under operating conditions, the best one can hope for is to prevent about 75% of the females from giving birth.

Other experts call for the removal of artificial water sources, which are sometimes used to lure the animals to within sight of tourists. In Addo, one can have lunch at the park restaurant and then stroll across the car park to a viewing area overlooking a water-hole, and

expect to see at least one or two massive grey beasts having a drink.

But there are at least two problems with removing the water-holes. At Addo, water-holes had to be put in because the river, which would have been the water source for elephants in the area, is not part of the park. Without the water provided by SANParks, the Addo elephants would perish along with all the other mammals. Another problem is that when you use water scarcity to limit elephant numbers, they die of thirst. This would be an even worse public-relations disaster than shooting them, according to Shrader. “It would hit the evening news: skeletal elephants and their babies.”

More controversially, Rudi van Aarde of the University of Pretoria advocates knocking down all the fences you can and allowing natural metapopulation dynamics to manage elephant numbers. The idea is to get as close to ‘natural’ as possible. “It is when you put up hard fences and dig the land full of water-holes, that we are creating problems that we shouldn’t try to solve through the barrel of the gun or through contraception,” says van Aarde.

Critics note that many people would be angry, not to say frightened, if the fences between their village and a herd of elephants were removed. Van Aarde insists that enough unpopulated land could be cobbled together to

avoid this, and that if elephants caused problems at the extremes of their range then they could be shot by villagers. But this would put the onus of management on local people with a lot to lose. “Maybe with time this will prove too extreme to be applicable,” he admits.

The argument over elephant ranges ultimately depends on how you draw the maps. Many people feel that the areas currently occupied by elephants in Africa are as large as can be reasonably expected. With human numbers growing as they are, there is no space for the elephants to expand into.

And even when range expansion has been tried, there’s scant evidence that taking down the fences encourages elephants to populate a new area. In many cases, they like it where they are. If they can get enough water, why move?

The removal of a fence on one side of Kruger to create a park that crosses into Mozambique has reportedly boosted the transit of people in stolen cars much more than it has increased the movement of elephants.

Packing their trunks

There are also proposals to move elephants from areas where they are living in high densities to areas where numbers are fewer, but according to Dublin, interest in that idea has “cooled down” in recent years. Moving an elephant is — no surprise here — an enormous undertaking. One must knock them out and then heft them around in trucks chaperoned by experienced personnel. It is very expensive. Another concern, according to Dublin, is the destination. For example, if it is sparsely populated because poachers are killing elephants there, why move

“I suspect that in the end we will come up with a uniquely African solution.”
— Graham Kerley



Effects of drinking: elephants can dramatically alter the landscape around their water sources.

K. GOUGH



Pushing for research: Graham Kerley works with officials at Addo Elephant National Park, which contains more elephants per square kilometre than some experts estimate it can sustainably support.

the beasts into a death trap? Some non-governmental groups such as the Massachusetts-based International Fund for Animal Welfare and the South Africa-based Peace Parks Foundation have funded relocations, but usually on the scale of tens of animals. The Kenya Wildlife Service, which manages elephants in that country's protected areas, relies heavily on relocations, as it does not cull.

Ask any elephant expert about the various management options, and they will usually say they may need to use all or some, depending on the situation. "You cannot expect that one size fits all," says Dublin. "If you take away any of the tools, that is not the best thing for elephants or for Africa."

Above all, elephant managers — be they governments, park managers, or consulting scientists — have to decide the reasons for managing elephant numbers. If it is to achieve maximum biodiversity, one strategy and population density of elephants might be best. If it is for iconic savannah landscapes, or to increase the genetic diversity of elephant populations, other routes may be needed. As Kerley says, "if you just want an elephant viewing park, you might as well just tear all this out and plant alfalfa".

Part of the challenge is that we have only vague ideas of elephant numbers and movements in Africa prior to Europeans showing up with flintlock elephant guns and sailing away with holds full of ivory. Some elephants migrate. Females and their offspring move about together; bulls often live alone. It is no easy task to establish a baseline — whether for elephant numbers or behaviour, or for the landscapes in which they lived.

Even the parts of Addo park that are

untouched by elephants are not a good guide to what the forests looked like before European colonization, because there would have been some elephants around. "What we don't have information on," says Kerley, "is what the landscape should look like and how many elephants will achieve that."

Some argue that culling, although the most controversial intervention, may also be the most 'natural'. Those who study elephants say that they do not change their reproductive behaviour much in the face of food scarcity — although van Aarde and others maintain that elephants do this in savannah ecosystems. Other herbivores will have their first offspring later in life and have subsequent offspring farther apart in lean times. Not so the exuberant elephant, which, given enough foliage to start with, will reproduce and consume its way right into a nasty population crash. So perhaps something else was limiting their numbers in ages gone by.

Mammoth story

We do know that before the Europeans arrived, Africans hunted elephants in pit traps, or, in what must have been a pretty spectacular manoeuvre, running up and slashing them in the Achilles tendon. David Cumming of the Percy FitzPatrick Institute of African Ornithology at the University of Cape Town speculates that human predation has been the key limit

on the population of elephant-like animals — proboscideans — in Africa for more than a million years. He is of the camp that favours predation over climate change as an explanation for the extinction of earlier proboscidean species.

The fossil record confirms that the number of proboscidean species dropped from nine to two after humans appeared on the evolutionary scene. Cumming mentions, too, the many large mammals that humans almost certainly did kill to extinction. So can South Africa's elephants prosper without culling? Is it possible that by protecting elephants in special areas, we are removing what once limited their numbers? Animal rights activists and tourists don't like hunting or culling. But it might be the simplest and most direct way to reduce elephant numbers.

Not all experts agree that pre-gun human predators were capable of taking down enough elephants to directly regulate the population. But even if it's not the 'natural order', it might still have a place in management.

And as Kerley explains, hunting isn't as simple as reducing the total number. "It's not how many elephants you kill, but how you influence resource use through fear. Elephants are incredibly risk averse." So by killing a few elephants near a certain village or farm or water-hole, you may be able to keep the rest of the herd, especially the more timid females and calves, from using that area.

Kerley argues for exploring all options: "I suspect that in the end we will come up with a uniquely African solution.

We might reinstate some level of predation — call it predation rather than culling — but we will also make more space for elephants." And the inescapable conclusion from all across Africa is that elephants do have to be actively managed. It is

interesting to look into their large dark eyes — famous for seeming wise, even world-weary — and wonder if they have any inkling that they are no longer in charge. ■

Emma Marris is a reporter for *Nature* based in Columbia, Missouri.

1. Kerley, G. I. H. & Landman, M. S. *Afr. J. Sci.* **102**, 395–402 (2006).
2. Johnson, C. F., Cowling, R. M. & Phillipson, P. B. *Biodivers. Conserv.* **8**, 1447–1456 (1999).
3. Slotow, R., Balfour, D. & Howison, O. *Pachyderm* **31**, 14–20 (2001).
4. IUCN African Elephant Status Report, www.iucn.org/themes/ssc/sgs/afesg/aed/pdfs/aesr2007.pdf

F. DU PLESSIS

K. GOUGH

Scientists should unite against threat from religion

SIR — It was genuinely alarming to encounter Ziauddin Sardar's whitewash of Islam in the pages of your journal ('Beyond the troubled relationship' *Nature* **448**, 131–133; 2007). Here, as elsewhere, *Nature's* coverage of religion has been unfailingly tactful — to the point of obscurantism.

In his Commentary, Sardar seems to accept, at face value, the claim that Islam constitutes an "intrinsically rational world view". Perhaps there are occasions where public intellectuals must proclaim the teachings of Islam to be perfectly in harmony with scientific naturalism. But let us not do so, just yet, in the world's foremost scientific journal.

Under the basic teachings of Islam, the Koran cannot be challenged or contradicted, being the perfect word of the creator of the Universe. To speak of the compatibility of science and Islam in 2007 is rather like speaking of the compatibility of science and Christianity in the year 1633, just as Galileo was being forced, under threat of death, to recant his understanding of the Earth's motion.

An Editorial announcing the publication of Francis Collins's book, *The Language of God* ('Building bridges' *Nature* **442**, 110; 2006) represents another instance of high-minded squeamishness in addressing the incompatibility of faith and reason. *Nature* praises Collins, a devout Christian, for engaging "with people of faith to explore how science — both in its mode of thought and its results — is consistent with their religious beliefs".

But here is Collins on how he, as a scientist, finally became convinced of the divinity of Jesus Christ: "On a beautiful fall day, as I was hiking in the Cascade Mountains... the majesty and beauty of God's creation overwhelmed my resistance. As I rounded a corner and saw a beautiful and unexpected frozen waterfall, hundreds of feet high, I knew the search was over. The next morning, I knelt in the dewy grass as the sun rose and surrendered to Jesus Christ."

What does the "mode of thought" displayed by Collins have in common with science? *The Language of God* should have sparked gasping outrage from the editors at *Nature*. Instead, they deemed Collins's efforts "moving" and "laudable", commending him for building a "bridge across the social and intellectual divide that exists between most of US academia and the so-called heartlands."

At a time when Muslim doctors and engineers stand accused of attempting atrocities in the expectation of supernatural reward, when the Catholic Church still preaches the sinfulness of condom use in villages devastated by AIDS, when the

president of the United States repeatedly vetoes the most promising medical research for religious reasons, much depends on the scientific community presenting a united front against the forces of unreason.

There are bridges and there are gangplanks, and it is the business of journals such as *Nature* to know the difference.

Sam Harris

Address withheld by request
www.samharris.org

Religion: Islamic science fading before colonialism

SIR — In his Commentary on the rise and fall of Islamic science (*Nature* **448**, 131–133; 2007), Ziauddin Sardar's discussion of how modern Muslim societies could be improved by scientific endeavours is encouraging. However, he places most of the blame for the decline of Islamic science on Western colonialism. This is historically inaccurate. Like the tip of an iceberg, colonialism is not the major factor — just the most visible.

It is true that colonialism "displaced meaningful cultural activities from Muslim society," and caused "the general economic and political deterioration of Muslim society" that led to the ultimate collapse of Islamic science. But it's generally accepted that the golden age of Islamic science stretched from about 800–1400 AD, and its decline started more than a century before Western colonialism began in the late fifteenth century.

A number of devastating blows to Muslim society have been implicated in the decline of Islamic science: military invasions, massacres and infrastructure destruction; a long period of drought beginning around 1250 AD; and a series of plague epidemics between 1347 and 1515. These destructive developments overlap with the main period of decline of Islamic science in the fourteenth and fifteenth centuries. The onset of European colonialism from the late fifteenth century onward merely completed a process that had begun long before.

Todd P. Silverstein

Department of Chemistry, Willamette University,
Salem, Oregon 97302, USA

Berlin shows how natural history can pull the crowds

SIR — You emphasized the roles that museums should play in both increasing and diffusing knowledge, in your Editorial 'Museums need two cultures' (*Nature* **446**, 583; 2007). Germany's premier natural-history institution, the Berlin museum of natural history, recently demonstrated the

potential that museums have to do the latter.

On 13 July, its main palaeontology exhibition reopened after a redesign. The dinosaurs in the main hall are now all mounted according to the latest findings in vertebrate palaeontology. The tallest mounted dinosaur skeleton in the world — of *Brachiosaurus brancai* — is no longer sprawling but striding majestically through the hall. On the opening day, thousands of visitors queued from the underground station, breaking Germany's record for the highest number of visitors to an exhibition during an opening weekend.

The Berlin *Brachiosaurus* skeleton was first mounted in 1937. Seventy years, although little by dinosaur standards, is a long time for scientists and the public to wait before money is provided for a modern exhibition. Funding bodies must recognize that natural-history museums are in a unique position to explain complex issues in science to a wide audience. But to do so, they need to be able to update their exhibits and present new research to the public on a regular basis.

The results are well worth the investment. During the opening weekend, more than 38,000 visitors preferred the Jurassic period to a day in the Berlin sunshine.

Nizar Ibrahim

School of Biology and Environmental Science,
University College Dublin, Belfield,
Dublin 4, Ireland

Puns can be baffling, so keep headlines simple

SIR — I beg to differ with Renée M. Ned and Lisa N. Steele's Correspondence 'Slang's not so slick when you remember its origins' (*Nature* **447**, 775; 2007) about the use of the word 'pimp' in a News Feature headline ('Pimp my antibody' *Nature* **446**, 964–966; 2007).

The word first appeared in sixteenth-century France as the verb *pimper*, meaning 'to dress elegantly', and as the adjective *pimpant*, 'alluring in dress, seductive'. In the seventeenth century, the word was associated with 'a knave, rascal, varlet, scoundrel', according to the Online Etymology Dictionary (www.etymonline.com). The vulgar modern meaning probably derives from a combination of these. The sense in which it is used by the television show *Pimp My Ride* could imply dressing an automobile elegantly — admittedly with a hint of flashy style.

I personally find the use of puns, colloquialisms and cultural references more objectionable, as they are likely to be understandable to only a fraction of *Nature's* global readership. The English language is sufficiently complex without the need to understand these sometimes obscure

references in the headlines of *Nature News* stories and other similar articles.

A quick scan of a few issues yields: "...over a pork barrel"; "Oceanography: Churn, churn, churn"; "...science in premier league"; "State of the donation"; "Astrophysics: The answer is blowing in the wind"; "Scot on the rocks"; and "The silence of the robins". As a native English speaker I may understand and appreciate these, but many others wouldn't.

Jeff Craig

Epigenetics Research Laboratory,
The Murdoch Childrens Research Institute,
Royal Children's Hospital, Flemington Road,
Parkville, Victoria 3052, Australia

Puns: wimp or macho, not a particle of offence is meant

SIR — Beyond the excellence of the scientific reporting, I particularly enjoy the entertaining use of language and the enlightened levity of *Nature*. Consequently, I am somewhat concerned by the complaint of R. M. Ned and L. N. Steele ('Slang's not so slick when you remember its origins' *Nature* 447, 775; 2007) regarding the use of the verb 'pimp', because of its 'immoral origins'.

Should I take offence because the use of the terms 'wimp' and 'macho' to denote putative particles might perpetuate negative stereotyping of my gender?

The freedom and, especially, the humour of scientific reporting may be hindered by misguided attempts to avoid offending moralists.

Milan Hopkins

9479 Main Street, PO Box 638,
Upper Lake, California 95485, USA

Regions unite to challenge inequalities in Brazil

SIR — The "unequal struggle" of medium-sized university physics departments in Scotland "against larger and more-entrenched rivals" is highlighted by your Editorial 'All for one...' (*Nature* 447, 1031; 2007). The less developed countries of the world face an even greater problem. How can science help the economic improvement of these countries, taking into account not only "basic challenges in trying to compete globally", but also regional difficulties — particularly very limited facilities and insufficient human resources?

As an illustration of the regional problem, take the Human Development Index, which measures quality of life: a score above 0.5 is medium and above 0.8 is high. The south

of Brazil scores about 0.84 — not far behind some European countries — whereas the northeast of Brazil is around 0.67.

Along the same lines as the Scottish Universities Physics Alliance described in your Editorial, the Brazilian government and universities created the Northeast Network on Biotechnology, or Renorbio (www.renorbio.org.br) in 2003. Last year a graduate programme in biotechnology was added, consisting of a pool of 29 universities from 10 different Brazilian states.

This initiative is new in the following five ways. It aims to produce 100 PhD graduates a year, which should speed regional development. Many universities that were not strong enough previously to offer a PhD programme now can, and will also gain research funds. Researchers who join the programme qualify for PhD students. The universities are becoming more competitive both locally and internationally. They are now expected to produce more patents and publications, creating stronger departments and leading to an increase in the number of knowledge-based enterprises in the region.

Luiz A. B. Castro*, **Allan Kardec Barros†**

*R&D Policies and Programmes,
Ministry of Science and Technology,
Esplanada dos Ministérios, Bloco E, Brasília,
DF, 70067-900, Brazil

†Department of Electrical Engineering, Federal
University of Maranhão, Av. dos Portugueses, s/n,
São Luís, MA, 65080-040, Brazil

Summing up *The Simpsons*

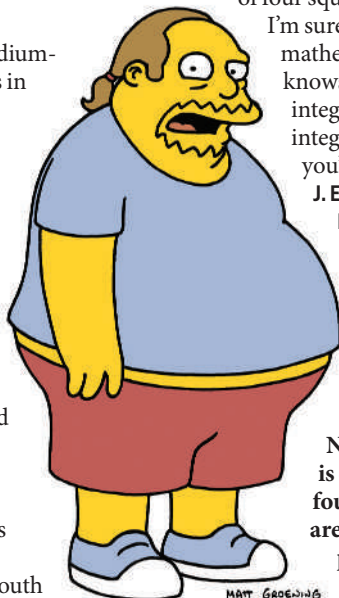
SIR — In your amusing News Feature interview with Harvard mathematics graduate Al Jean, head writer of *The Simpsons* ('Mmm... Pi' *Nature* 448, 404; 2007), Jean mentions, as an example of a staggeringly obscure mathematical reference, a number in the thousands that is the sum

of four squares. However, as I'm sure every Harvard mathematics graduate knows, every positive integer is the sum of four integer squares. I think you've been had!

J. Ewart H. Shaw

Department of Statistics,
University of Warwick,
Coventry CV4 7AL, UK

The true significance of 8,208, the number referred to in the News Feature, is that it is one of only three four-digit numbers that are the sum of the fourth powers of their digits — Editor, *Nature*.



Drop 'higher' and 'lower' to raise descriptive standards

SIR — For most of the past 2,500 years, the dominant view in the West was that life was graded from better to worse, higher to lower, in a Chain of Being. This could be the product of some creative deity or — as in the eighteenth- and nineteenth-century theories proposed by Denis Diderot, Jean Baptiste de Lamarck and Robert Chambers — a consequence of matter somehow organizing itself through a goal-directed, progressive evolutionary process. Humans invariably occupied the top link.

The scientific literature persists in providing a home for the terminology of the Chain. During the past year, for instance, the Nature.com search facility identifies more than 300 references to 'higher' or 'lower' eukaryotes or organisms.

This usage is misleading, because evolution is not a progressive process. Selection may, in the short term, lead to increases in fitness. But there is no reason to assume that species adapted to the environments they occupy today are better adapted than their precursors were to their environments, and if evolution does not ensure this it hardly qualifies as progressive. Of course, because the first life forms were very simple, it is not surprising that evolution has resulted in increases in specialization or complexity in many lineages. But there is no sense in which such changes can make one taxon higher than another.

An argument in favour of preserving the terminology of the Chain is that it is useful. Yet there is considerable disagreement over the referents of 'higher' and 'lower'. The status of plants, for example, as higher or lower organisms is very fluid. Yeast is typically a 'lower' organism when compared with animals but can be a 'higher' one when compared with bacteria. Perhaps more disturbingly, and in a usage reminiscent of natural theology, non-human mammals can be relegated to the group of lower organisms when they are compared with humans.

A descriptive terminology hardly qualifies as useful if users disagree over what it describes. Moreover, it is rarely the case that using a consistent and scientifically robust terminology is more difficult than using the inconsistent and misleading terminology of the Chain. Distinguishing between plants and animals, for example, or between flowering and non-flowering plants, or between vertebrates and invertebrates, or between humans and other primates/mammals/animals without resorting to the descriptors 'higher' and 'lower' is easy — it's just been done.

Michael Mogie

Centre for Mathematical Biology,
Department of Biology and Biochemistry,
University of Bath, Bath BA2 7AY, UK

BOOKS & ARTS

Four ways to take the policy plunge

How should researchers best interact with policy-makers for maximum benefit to society?

The Honest Broker: Making Sense of Science in Policy and Politics

by Roger Pielke Jr

Cambridge University Press: 2007. 198 pp.
\$29.99, £15.99

Andrew A. Rosenberg

Few scientists are trained to be policy advisers. Many academics look askance at politics, as though there might be something unseemly about bringing their research to bear on problems of the day. This is probably not so much from disinterest as from a lack of experience in real-world policy debates and of an academic incentive to participate in them. Happily, the book by Roger Pielke Jr on the engagement of scientists in policy offers a pithy, insightful basis for discussing the contributions scientists can make to advising policy-makers.

Pielke sets out four basic modes for how scientists advise on policy and two models for scientific advice. The first mode is for the 'pure scientist', who is not involved at all. This is the colleague who shakes their head ruefully because the policy-makers were foolish enough not to read his or her recent papers. The second is for the 'science arbiter', who readily responds to questions but without voicing an opinion on the policy choices. Again, this person sets themselves apart from the problems of the policy-maker and is loath to take ownership of the results.

The third role is the 'issue advocate', who takes a clear position and argues for specific policy action, using scientific knowledge and training to make that argument. Finally, the 'honest broker' — Pielke's obvious preference — engages in the policy process to bring scientific information to bear by considering the options and helping policy-makers to choose. Here, there is real ownership of the process, and products of policy as options are explored from a scientific perspective.

These caricatures of the roles that science advisers play provide useful scope for thought about how to interact with policy-makers. I have had experience with all four roles at various points in my career as both scientist and policy-maker, and agree with Pielke that scientists often make the most impact by serving as honest brokers. But his view of issue advocates is excessively disparaging. Although a scientist who takes a strong position on policy may not comport with the view of some that science is somehow 'pure' and above the messy fray of



At the 2007 Intergovernmental Panel on Climate Change conference, scientists help shape policies.

policy debates, such advocacy has had a hugely positive effect on policy.

Issue advocates are to be welcomed and the scientific evidence weighed as it always should be, irrespective of who brings it to the table. It is too easy to label a scientist as an issue advocate to downplay their opinion without a rigorous reason. In my experience, most scientists who take the role of issue advocate do so out of conviction resulting from their work, in much the same way as researchers will argue strongly for their interpretation of research results. It is therefore inappropriate to assume that issue advocates rely on science that is less rigorous.

The strongest case that Pielke makes is this: "That some scientists engage in political activities is neither new nor problematic; they are after all citizens. A problem exists when ... scientists implicitly or explicitly equate scientific arguments with political arguments, and in the process reinforce a simplistic and misleading view of how science supports policy." The challenge is to separate carefully which of your inferences stem solely from the science, and which from your views as a citizen.

The two models for science advice Pielke describes are the linear and the stakeholder models. The first is a simple progression from science to policy, with no real feedback; the second is an integrated system. I tired of Pielke railing against the linear model, and more so of

the implication that this model somehow leads to scientists unwittingly becoming issue advocates instead of honest brokers. Stakeholder involvement is essential, but can accentuate the tendency to polarize positions and for scientists to advocate one policy over another.

Pielke's choice of the basis for the Iraq war as an example of advisory processes, and how they fail, is a missed opportunity. An alternative scientific case study would have been more relevant to the role that most scientists may play in policy-making. For example, the natural-resource issues that I have dealt with in fisheries and the marine environment have extensive advisory processes and stakeholder input. Scientists were engaged in these issues in all of Pielke's four roles. The process is difficult and time-consuming, but it is likely to be much more relevant to the work of most scientists than the gathering of military intelligence.

This is a clear, thought-provoking book that helps move us away from thinking of science as 'pure' and distinct from policy. It would make an excellent basis for a graduate seminar. It isn't a textbook, but a think-piece, and we all need to consider carefully our responsibility to engage as scientists in policy-making. ■

Andrew A. Rosenberg is professor of natural resources at the Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, New Hampshire 03824, USA.

J. BRINON/REUTERS

Mankind's strange love of superweapons

Doomsday Men: The Real Dr Strangelove and the Dream of the Superweapon

by P. D. Smith

Allen Lane/Penguin: May 2007.

553 pp. with index, £20

St Martin's Press: December 2007. 416 pp.

\$27.95

Gregg Herken

"There is nothing in Man's industrial machinery but greed and sloth: his heart is in his weapons," said the Devil in George Bernard Shaw's *Man and Superman*. Shaw's adage could almost be the *leitmotiv* of P. D. Smith's well-researched and altogether depressing account of humankind's long hunt for the ultimate superweapon: a doomsday device that, by its very terribleness, would make war forevermore unwinnable, and hence unthinkable. Although we all know how this tale turns out, it is a journey well worth taking. Along the way, Smith includes some fascinating asides about the men — and it was, almost exclusively, a fraternity — who, in seeking to make war obsolete, have only made it more deadly.

Typical of the breed was German chemist Fritz Haber, who invented and personally promoted the use of poison gas on First World War battlefields. As with his contemporaries in the other belligerents, Haber started out as an idealist who, once hostilities began, quickly harnessed his science in the service of the state. His motto, "*Im Frieden der Menschheit, im Kriege dem Vaterland*" (In peace, for mankind; in war, for the fatherland), sets a pattern followed with distressing regularity by other idealistic scientists in subsequent conflicts.

The real focus of Smith's book is Leo Szilard, a Hungarian-born American physicist who fled Nazi Germany and eventually wound up in the United States. Szilard was arguably the first to envisage how a nuclear weapon might be made. "You know what fission means," he confided to fellow physicist Edward Teller in early 1939, "it means bombs."

Ironically, it was a seemingly off-hand remark made by Szilard in a nationwide radio broadcast in February 1950 — shortly after the Truman administration announced plans to proceed to the next level of destructiveness, the hydrogen bomb — that revived interest in a true 'doomsday' weapon. Szilard envisaged building a battery of very powerful hydrogen bombs jacketed with cobalt-60, an element with a radioactive half-life of 5.7 years. In a later iteration, the bombs would be hooked up to a computer, which, on receiving a certain signal — say, a single nuclear weapon exploding on friendly territory, or the crossing of a border by enemy troops — would detonate them. Theoretically, such a device, if set off, would enshroud Earth with lingering, lethal levels of radioactivity, snuffing out all human life.

In the end, the cobalt bomb was never very



Leo Szilard was among the first to conceive of a nuclear weapon.

attractive to even the most bloody-minded of strategists because it would not have been destructive enough: the longer the half-life, the less intense the radiation. That did not stop a generation of novelists and film-makers from enlisting the cobalt bomb in their doomsday scenarios: among them, novelist Nevil Shute in *On the Beach* and director Stanley Kubrick in *Dr. Strangelove*. Nor did it prevent P. D. Smith from making the cobalt bomb the

centrepiece of his book.

Doomsday Men is prodigiously researched, the author seeming to have read everything on nuclear strategy, both fiction and non-fiction (admittedly, the two are sometimes difficult to tell apart). However, there are some nagging deficiencies. Editing is one. We are told twice within a few pages that Leo Szilard was considered "the greatest scientist never to have won a Nobel Prize." Focus is another. The book is as much a history of modern science as of modern weaponry, so dominant has been the latter in the evolution of the former. But it is not clear, for example, what connection the discovery of X-rays has with the advent of doomsday weapons.

One can only sympathize with the author's observation that, since the end of the Cold War, global warming and Islamist terrorism have distracted our attention from the weapons that

remain in the arsenals of nations, numerous, primed and waiting. Although not as deadly as Smith's fictive doomsday bomb, they are cause for us to be more fearful, for they are real. ■

Gregg Herken is professor of history at the University of California at Merced, PO Box 2039, Merced, California 95344, USA, and author of *Brotherhood of the Bomb: The Tangled Lives and Loyalties of Robert Oppenheimer, Ernest Lawrence, and Edward Teller*.

Linnaeus lives on

Order out of Chaos: Linnaean Plant Names and their Types

by C. Jarvis

The Linnean Society of London: 2007.

1,016 pp. £80

Pamela S. Soltis

The relevance of taxonomy in our genomic era is greater than ever. Correct naming is crucial for developing new foods and medicines, and for understanding our changing environment. Amazingly, we do not even know how many species of flowering plant currently exist. Estimates range from about 250,000 to more than 400,000, and every day species are lost to human activities. Conservation, in natural habitats or botanical gardens, relies on scientific names, as does the rest of science.

Before Carl Linnaeus, species names ranged from cumbersome to unwieldy. For exam-



ple, *Arbutus caule erecto, foliis glabris serratis, baccis polyspermis* was as good as it got for 'Arbutus with upright stems, hairless, saw-toothed leaves, and many-seeded berries'. Linnaeus's binomial

system of nomenclature changed all that. Species could now have a genus name plus a specific epithet. 'The father of taxonomy' applied his revolutionary system to plants in *Species Plantarum*, published in 1753, and to animals in the tenth edition of *Systema Naturae* in 1758. More than 250 years later, the name of a species remains the key that unlocks all that is known about it.

In honour of Linnaeus's 300th birthday, Charlie Jarvis's handsome new book, *Order out of Chaos*, gives us a glimpse into pre-linnaean botany, Linnaeus's life and work,

and the reception and impact of his ideas in eighteenth-century Europe.

Linnaeus was an avid collector and classifier of 'animals, vegetables and minerals'. He named and described his collections and the many specimens that were returned to Europe from exploration of other parts of the globe. His plant names are the starting point for valid names that follow botanical nomenclature. To be valid, a species name must be attached to one reference or 'type' specimen. This concept was implemented 150 years after *Species Plantarum*, so — ironically — most of the names assigned by Linnaeus himself do not have clearly associated specimens.

Enter the Linnaean Plant Name Typification Project, launched in 1981 at the London Natural History Museum to match Linnaeus's thousands of designations with the specimens and illustrations in his work. Jarvis's book chronicles this project and presents nearly 700 pages of names and their types — from *Abrus precatorius* L. to *Zygophyllum spinosum* L. Jarvis and his colleagues scoured museums and private collections to create this wonderful



One of many plant species named by Linnaeus.

reference for all those involved in taxonomy.

The book is a work of art. Even non-botanists will be seduced by this inviting package. The text is peppered with photographs and illustrations of spectacular examples of plant diversity and linnaean memorabilia. Figures range from a photograph of Linnaeus's signature and seal from 1734, to an illustration of the wild woodland twinflower (*Linnaea borealis*), to Georg Dionysius Ehret's (1736) illustration of Linnaeus's 'Sexual System' of plant classification.

The appreciation for history in plant systematics is evident in this book and in the attention being paid to Linnaeus during 2007. Even the controversy between those who favour the continued use of linnaean classification and advocates of the PhyloCode (the system that relies on evolutionary history) does not diminish Linnaeus's contribution. This book is a timely homage to linnaean names and the efforts of Jarvis's team, as systematists look for new ways to provide the richest classification systems. ■

Pamela S. Soltis is at the Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611, USA.

The theatre of quantum physics

Faust in Copenhagen: A Struggle for the Soul of Physics

by Gino Segre

Viking/Jonathan Cape: 2007. 320 pp.
\$25.95/£20.00

Finn Aaserud

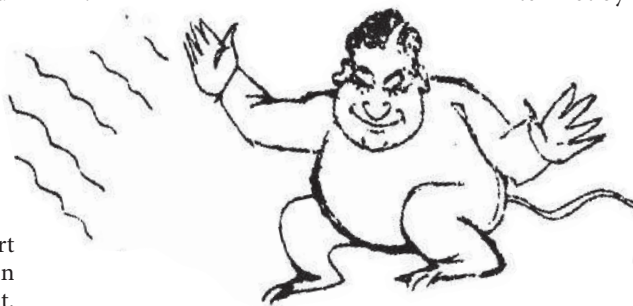
The year 1932 was a particularly eventful one in physics. In *Faust in Copenhagen*, the physicist Gino Segre chooses as his point of departure an annual meeting held in April of that year at Niels Bohr's Institute for Theoretical Physics at the University of Copenhagen, Denmark. In Segre's view, the 1932 gathering symbolized the end of the political neutrality of physics and physicists, coinciding as it did with the arrival of Hitler and crucial discoveries in nuclear physics that would make possible the subsequent development of the atomic bomb.

Segre introduces us at the outset to seven physicists who attended this series of informal but prestigious meetings: Niels Bohr, Paul Dirac, Werner Heisenberg, Paul Ehrenfest, Max Delbrück, Lise Meitner and Wolfgang Pauli. He presents a fascinating comparison of these main characters in the first hundred pages, and then launches into a spirited and engaging history of the development of quantum physics from about 1900 to 1932. Segre's emphasis is always on the human aspect, and he is unable in this part of the book to limit his story to the seven physicists whom he claims the book is about. Towards the end, he returns to the original group, vividly describing their subsequent

careers, again with illuminating comparisons.

As the nephew of Emilio Segrè, a Nobel prize-winning physicist trained by some of the main characters in the book, Segre deploys his background effectively for dramatic and comparative purposes. This, however, shortens his distance from the subject matter — for example, the role of experimental physics at Bohr's institute is underplayed. Moreover, he seemingly contradicts his interpretation of the 1932 meeting as the end of the political neutrality of physics, by praising his own field for continuing to this day in the disinterested tradition instigated by Bohr.

Bohr's crucial contributions to quantum mechanics, Segre writes, were "largely ignored" by Segre's own generation of physicists, and he seems intent on rectifying the situation. Whether or not this is necessary, Segre's approach reveals important insight into Bohr as a person as well as into his interaction with and importance for physicists who followed him.



Wolfgang Pauli as the Devil in a parody of *Faust*, performed by physicists at a meeting in 1932.

At the end of the 1932 meeting, in celebration of the hundredth anniversary of Johann Wolfgang von Goethe's death, some of the younger scientists performed a parody of *Faust*, replacing Goethe's original characters with some of the main theoretical physicists of the day, most of whom were also assembled in Copenhagen: Niels Bohr was portrayed as God, Wolfgang Pauli as the Devil and Paul Ehrenfest as Faust. The parody described the state of physics at the time, and Segre regards it as sufficiently important to name his book after it.

At least to me, however, Segre does not integrate the play into his narrative satisfactorily. The worldwide success of Michael Frayn's play *Copenhagen*, first set up in London in 1998, inspired a still-active interest in the relationship between science and the theatre. Frayn's play even shares two of Segre's main characters — Niels Bohr and Werner Heisenberg. But *Faust* as enacted at the 1932 meeting was written not by a



professional playwright but by the participant physicists themselves, and one might have hoped that Segrè would have used the opportunity to offer a new twist to the relationship in question. Admittedly, there are quotations from the original *Faust* at the head of each chapter, as well as a chapter entitled 'Goethe and Faust' and another comparing the physicists' play with Goethe's original. But these instances are provided largely in isolation from the main narrative of the book.

In 1932, German was still the main language

of physics — and culture — so the choice of Goethe's *Faust* as a basis for a satire enacted in German seems natural. Probably in anticipation of his likely readership, Segrè does not quote from the original German, but quotes instead from necessarily inferior English translations. The many quotations from Goethe's *Faust* are taken from Anna Swanwick's excellent, but by now somewhat archaic, translation from late in the nineteenth century. Likewise, Segrè's treatment of the physicists' *Faust* is based on the spirited, but sometimes inaccurate,

English translation published in 1966 in a highly personal account, *Thirty Years That Shook Physics*, by the physicist George Gamow (to whom Segrè dedicates his book).

The photographs used to illustrate *Faust in Copenhagen* are well chosen. Apart from a few minor omissions and factual errors, the book is lively and readable and provides an exciting impression of the development of theoretical physics during a crucial period.

Finn Aaserud is director of the Niels Bohr Archive, Blegdamsvej 17, DK-2100 Copenhagen, Denmark.

Medical history without frontiers

Medieval Islamic Medicine

by Peter Pormann and Emilie Savage-Smith

Georgetown University Press/Edinburgh University Press: 2007. 223/256 pp. \$29.95/£18.99

Yasmin Khan

What is 'Islamic medicine'? It is a term fraught with limitations. No definitive definition exists, but most reputable historians of medicine have characterized it by several key features. These include a chronological marker that spans a medieval period of a thousand years and encompasses specific but vast geographical regions from Spain and North Africa in the west, to central Asia and India in the east. Its successes were inspired implicitly by the Islamic faith, but sometimes involved discreet collaboration of Muslims with non-Muslims, who used Arabic as the *lingua franca* to cultivate the resulting phenomenon.

These idiosyncratic dynamics of Islamic medicine are explored in *Medieval Islamic Medicine*. With their meticulous research on newly translated manuscripts, two pre-eminent scholars of medical history, Pormann and Savage-Smith, have revisited the topic discussed in the classic text *Islamic Medicine* by Manfred Ullmann, published almost 30 years ago.

Their new, brief overview is destined to become a definitive entry-level textbook, owing in part to its simple structure, comprehensive analysis and succinct interpretation of key historical figures and their achievements. It relies on a thematic rather than chronological approach, which enhances its readability.

The book discusses the emergence of medieval Islamic medicine and the subsequent cross-pollination with medicine in other cultures. We are given a sense of how Muslims perceived the body as well as the soul as precious, because it was derived from and accountable to God as the creator. The body therefore required constant and dutiful care, and had the right to be maintained and preserved and protected from

abuse, meaning that the individual was bound to seek a cure for any ills. These sensibilities were influenced by precursor civilizations — particularly ancient Greece — and reinforced by the Islamic faith.

Pormann and Savage-Smith refute the clichéd view that the Islamic medieval civilization was merely a bridge between ancient and modern medicine. They show how Muslims built on Greek knowledge, in addition to preserving and translating it. The resulting new ideas and theories later percolated further into the West through translation, trade and travel, as well as



A Persian physician gives instructions for preparing smallpox remedies in the eleventh century.

during the crusades, acting as a springboard towards the European renaissance and thus forming the roots of modern medicine.

The book emphasizes the inclination of early Muslim medical practitioners towards *Unani Tibb*, a form of Islamic medical tradition stemming from early Greek medicine. This involved balancing, through diet and medicinal herbs, the four humours — air, earth, fire and water, which correspond to the four bodily fluids blood, phlegm, black bile and yellow bile — and understanding that a good environment and spiritual peace are essential for good health. Today such a holistic approach to well-being, although experiencing a resurgence, is outside mainstream modern medicine.

The authors avoid the irksome cumulative

and linear approach that besets many texts and museum displays of medical history, which often portray the past as inferior and the present as the benchmark of progress. And without taking the 'great man of history' line, they give us Islamic medical pioneers aplenty — for example, Abu al-Qasim al-Zahrawi (Albucasis). This accomplished surgeon from eleventh-century Muslim Spain first described and used forceps for extracting fetuses during stillbirths, developed other specialized surgical tools, and sutured wounds using cat gut.

Pormann and Savage-Smith also discuss the role of women in the development of medieval Islamic healthcare. Patronage by wealthy women associated with those in power proved instrumental in providing the financial investment needed for establishing hospitals in Baghdad from the ninth century. These institutions offered free, specialized treatments, which resembled the founding ethos of the modern British National Health Service.

What led to the demise of the stronghold of Islamic medicine in general is not addressed in the book: this is an unresolved and complex question. Some blame religion, or the lack of it, and others point the finger at social factors such as war and colonialism. This subsidence has been perpetuated and exacerbated by the 'brain drain' phenomenon. That said, the legacy created by the practice of Islamic medicine is alive and thriving in the rural Middle East, throughout Africa, among the Muslims of southern Asia and in Asian diaspora communities in Europe and the United States, indicating that there is scope for a wider revival.

Medieval Islamic Medicine is a measured and scholarly book that is a timely rebuke to those tempted to dismiss the past as irrelevant. The medieval period when the Islamic world drove medical development and discovery is an indication that Islam itself is not inimical to progress. This book should be read by anyone interested in the real story of civilization. It is a notable contribution to the promotion of a multicultural public understanding of the history of science.

Yasmin Khan is exhibition coordinator at the Science Museum, Exhibition Road, London SW7 2DD, UK.

NEWS & VIEWS

NEUROSCIENCE

Obsessed with grooming

Steven E. Hyman

Roughly 2% of humans suffer from obsessive compulsive disorder, but a lack of animal models has impeded research into this condition. Could a genetically engineered mouse model provide an exciting lead?

Obsessive compulsive disorder (OCD) is one of a spectrum of disorders characterized by obsessions (intrusive, unwanted thoughts) and compulsions (ritualized behaviours intended to overcome the anxiety and tension resulting from the obsessions). Other similar conditions include Tourette's syndrome; sufferers with this disorder show several motor, and occasionally vocal, tics, often accompanied by obsessions and compulsions. Although the evidence is less convincing, another disorder that might fit into this spectrum is trichotillomania, the hallmark of which is compulsive pulling out of scalp hair¹. On page 894 of this issue, Welch *et al.*² describe a genetically engineered mouse that shows some behavioural features similar to those of the obsessive compulsive spectrum of disorders in humans³.

Often beginning early in life and running a chronic and relapsing course, OCD causes significant distress and disability. The available treatments for this condition are only moderately effective. They include cognitive behavioural psychotherapies and antidepressant drugs that increase levels of the neurotransmitter serotonin at synapses (junctions between neurons). As is often the case in psychiatric disorders, the existing drugs for OCD stem from exploitation of serendipitous clinical observations. Neither the pathophysiology of the disorder nor the mechanism of action of these drugs is understood.

Given the toll of OCD and other psychiatric disorders, safer and more effective medications are much needed; but four main obstacles remain. First, understanding the neurobiology of higher cognition, emotion and control of complex behaviour is still a daunting frontier. Second, those with psychiatric disorders lack obvious and visible signs of damage to the nervous tissue, the presence of which could point the way to identifying molecular culprits. Third, the genetic and non-genetic factors contributing to psychiatric disorders are highly complex⁴. And fourth, good animal models have been lacking.

The mice studied by Welch *et al.*² showed excessive grooming, which resulted in hair loss and skin injuries, as well as anxiety-like traits. These mice lack the gene encoding SAPAP3



Compulsive hand-washing is a common symptom of obsessive compulsive disorder.

T. S. KENNEDY

— a scaffolding protein that is found in excitatory, glutamate-responsive synapses and is highly expressed only in the striatum region of the brain. The behavioural abnormalities in these mice were reversed by local expression of *Sapap3* in the striatal region, which indicates that loss of this gene is responsible for the observed behavioural abnormalities.

The authors also found that a drug from the SSRI class — which selectively enhance serotonin-mediated neurotransmission throughout the brain — that is used to treat OCD in humans decreases both grooming and anxiety in these mice. This is interesting because a condition responsive to an enhancer of serotonin neurotransmission does not signify a primary defect in serotonin-mediated signalling; instead, the defect is in glutamate-responsive synapses. So alterations in serotonin seem to modulate glutamate action. These findings are also noteworthy because Welch and colleagues have generated a possible mouse model of OCD. Moreover, these observations add to the accumulating, if circumstantial, evidence that OCD and its associated disorders result from abnormalities in neural circuits

spanning the frontal, striatal and thalamic regions of the brain.

Studies involving structural neuroimaging of the striatum in the brains of patients with OCD^{5,6} have often indicated abnormalities in the volume of striatum components, although they do not agree on the nature of these changes. However, a long-term study⁷ of patients with Tourette's syndrome suggests that the relationship between the volume of the caudate nucleus (a component of the striatum) and the symptoms of this disorder is complex; at the time of the scan (childhood), smaller caudate volumes did not correlate with symptoms, but they did predict persistence of severe symptoms into early adulthood.

Functional neuroimaging studies⁸ on patients with OCD, based on cognitive and emotional tasks, also point to abnormalities in frontal–striatal–thalamic circuits, although we cannot yet claim to be on firm ground. That said, a striatal origin for OCD-like symptoms would make a lot of sense. The medium spiny neurons of the striatum receive convergent glutamate-mediated inputs from the cerebral cortex, which provide detailed information

about the context in which behaviours occur. These neurons also receive dopamine-mediated inputs from the midbrain that report on the significance of the behaviour — for example, whether it yields an unexpected reward.

When something important happens, the striatum stimulates circuits that project via the thalamus to the prefrontal cortex; this leads to the formation of memories that can later guide the planning and control of behaviour. Abnormalities in synaptic function within frontal–striatal–thalamic loops could lead to unintended behaviours or even unintended thoughts. The main significance of the *Sapap3*-deficient mouse is that it represents a model of how such a pathological process might work. In this case, abnormal synaptic function causes grooming behaviours that are repetitive enough to produce skin injuries and unresponsive enough to control by the prefrontal cortex that they continue even after the injuries have occurred.

The mouse model described by Welsh and colleagues will excite those interested in the pathophysiology of OCD or in developing treatments for it. But we must retain a healthy caution. Genetic animal models of psychiatric disorders have been produced by inbreeding for desired traits⁹, by introducing human disease genes¹⁰, by mutating existing animal genes¹¹ or, as in this case, by observing an unexpected behavioural response in a genetically engineered animal. It is highly unlikely that such animal models will ever recapitulate human psychiatric disorders in their entirety. However, as in this case, it is possible to model significant aspects of these diseases.

The core symptom of OCD — unwanted intrusive thoughts — cannot be mimicked in mice, at least not in any obvious way. Instead, *Sapap3*-deficient mice show excessive grooming that continues past the point of self-harm. This is reminiscent, perhaps, of compulsive hand-washing, which in patients with OCD represents a never fully successful attempt to neutralize fears of contamination.

In OCD patients, the main cause of anxiety is the unwanted intrusive thoughts. The sufferers are anxious because they cannot be certain that the door is locked, the gas has been turned off, or that they are free of dreaded microbes. The anxiety-like behaviours observed in these mice may also resemble OCD, but this requires a stretch of the imagination.

Even if we can gain assurance with additional research that the behaviours observed in *Sapap3*-deficient mice reflect abnormalities in circuits that produce human symptoms, we cannot assume that OCD-related conditions in humans involve variations in this gene. These disorders, like other major psychiatric diseases, seem to be heterogeneous with complex underpinnings — probably involving several genes — that, in interaction with developmental and environmental factors, could lead to abnormalities in frontal–striatal–thalamic circuits.

Despite these reservations, the work of Welsh

*et al.*² sharpens our focus on frontal–striatal–thalamic circuits both in human patients and in animal models of OCD. It also gives us a compelling clue that the compulsive behaviour associated with this condition is due to a synaptic abnormality in these neural loops. Such cellular insight should aid an understanding of OCD at a molecular level. ■

Steven E. Hyman is in the Department of Neurobiology, Harvard Medical School, Massachusetts Hall, Cambridge, Massachusetts 02138, USA.

e-mail: seh@harvard.edu

QUANTUM PHYSICS

Wave goodbye

Luis A. Orozco

When measuring photons, it's a case of 'wanted, dead' — catching them alive is not an option. But we can observe how a superposition of many photon waves progressively collapses as it interacts with a beam of atoms.

Earlier this year, a team from the Ecole Normale Supérieure in Paris recorded jumps of light heralding the birth and death of a photon trapped in a cavity¹. As they describe in this issue (Guerlin *et al.*, page 889)², the same researchers have now performed a similar, more complex trick — recording exactly how a coherent state of many photons collapses as it is measured.

A measurement process differs fundamentally between the classical and quantum worlds. In the classical realm, there is no explicit limitation on a measurement's accuracy. In the quantum domain, by contrast, accuracy is constrained by the Heisenberg uncertainty principle: a measurement will produce a definite result, but one whose value is distributed according to the laws of probability. What is more, the measured object will itself be fundamentally altered by the measurement. Thus, the clicking sound produced when a photon is caught by a detector says two things: yes, a particle was detected; but sorry, the way you detected it killed it, and its energy was converted into an electric pulse.

But the quantum world has more subtle states to investigate than a single photon. Photons, or the probabilistic wavefunctions associated with them, can add together, or superpose. If they superpose coherently (in phase), their combined wavefunction begins to look like a classical wave. This coherent electromagnetic field is the complex beast whose collapse was monitored by Guerlin *et al.*².

But how did they achieve this feat, given the difficulties of measuring a quantum object without instantly destroying it? The authors' 'quantum non-demolition measurements' in a cavity quantum-electrodynamical (QED) system required profound understanding of quantum mechanics, continuous theoretical

1. Chamberlain, S. R., Menzies, L., Sahakian, B. J. & Fineberg, N. A. *Am. J. Psychiatry* **164**, 568–574 (2007).
2. Welch, J. M. *et al. Nature* **448**, 894–900 (2007).
3. Hollander, E., Kim, S., Khanna, S. & Pallanti, S. *CNS Spectr.* **12** (suppl. 3), 5–13 (2007).
4. Merikangas, K. R. & Risch, N. *Am. J. Psychiatry* **160**, 625–635 (2003).
5. Robinson, D. *et al. Arch. Gen. Psychiatry* **52**, 393–398 (1995).
6. Pujol, J. *et al. Arch. Gen. Psychiatry* **61**, 720–730 (2004).
7. Bloch, M. H., Leckman, J. F., Zhu, H. & Peterson, B. S. *Neurology* **65**, 1253–1258 (2005).
8. Remijnse, P. L. *et al. Arch. Gen. Psychiatry* **63**, 1225–1236 (2006).
9. Bice, P. J. *et al. Behav. Genet.* **36**, 248–260 (2006).
10. Shahbazian, M. *et al. Neuron* **35**, 243–254 (2002).
11. Clapcote, S. J. *et al. Neuron* **54**, 387–402 (2007).

elucidation of subtle details of cavity QED, and unprecedented dedication in realizing a simple theoretical model in the laboratory. This model³ first required the development of a pair of superconducting mirrors for the walls of the cavity whose losses are low enough that light remains captured between them for the length of time it would take the light to circle Earth at the Equator.

The second pivotal ingredient is individual rubidium atoms in a 'Rydberg' state in which one electron is highly excited. These atoms are like little planetary systems, with the excited electrons on a distant orbit around a remote atomic nucleus. They can oscillate between two different excited states, and the regularity of this oscillation makes them excellent time-keepers. The frequency of that oscillation is easily disturbed in the presence of light — to the extent that it can be used to detect the presence of a single photon non-destructively¹.

And so Guerlin *et al.*² prepared a coherent state of a microwave electromagnetic field in their cavity that contained up to seven photons, and then sent in the Rydberg atoms. They made sure that these atomic clocks had their 'hands' — their initial phase — set to one of eight possible values. Each of these settings is a possible answer to the question "How many photons are there in the cavity?" The interaction with the electromagnetic field modifies the setting of the hands according to the answer, from none to seven. The atoms leave the cavity, and the new direction of the hands is recorded.

Guerlin and colleagues' measurement method had to fulfil certain conditions for success⁴. First, the interaction of the quantum probe (the atom) with the remaining, classical part of the measuring device (the 'click' of the detector) must not begin until the

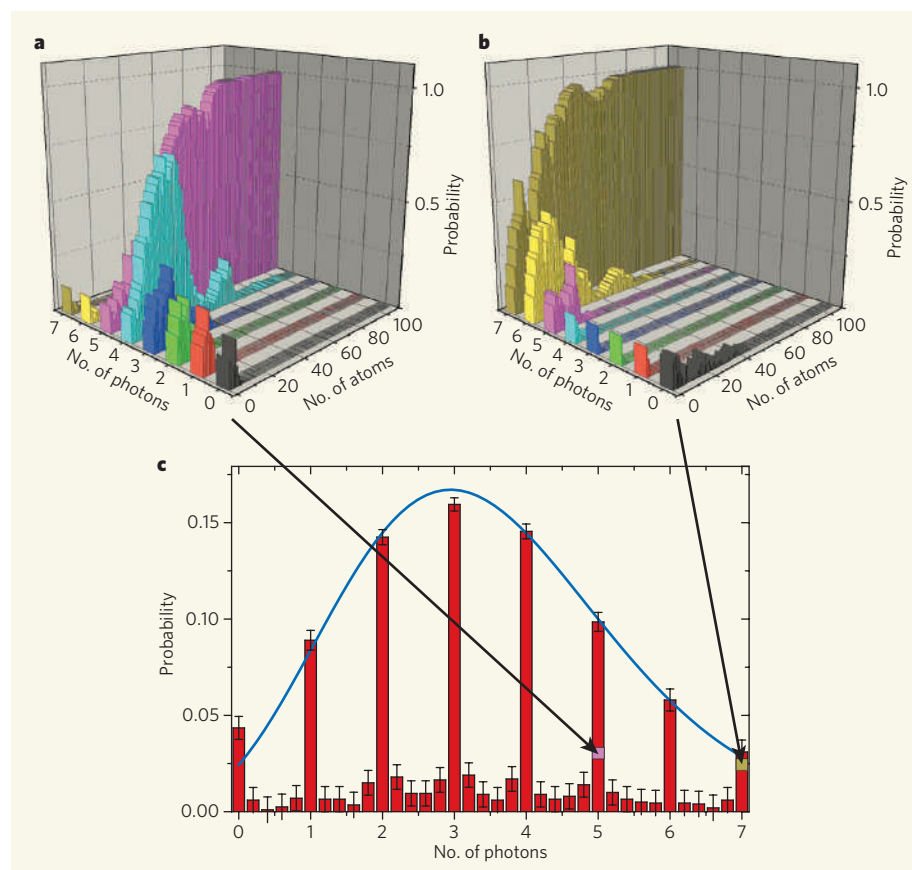


Figure 1 | State surveillance. **a**, As a stream of rubidium atoms passes through a mirror-walled cavity filled by a coherent light field in Guerlin and colleagues' experiment², the atomic 'clocks' are disturbed according to the number of photons present in the light field. The probability for different photon numbers is initially evenly distributed between the possible outcomes (none to seven photons), but 'collapses' to a definite value as more and more atoms pass through — to five, in this case. **b**, A different measurement of the same state collapses to a different photon number — seven. **c**, By repeating the same measurement over and over again, an overall (poissonian) probability distribution for the number of photons in the initial coherent light field can be gathered. (Figure modified from ref. 2.)

interaction of the atom with the electromagnetic field is complete. Second, the interaction with the classical part must not contribute significantly to the total error of the measurement. Finally, there should be no fundamental constraint on the precision of a measurement: in quantum-mechanical terms, the phase of the electromagnetic field (which according to the uncertainty principle is the 'canonical conjugate' of the number of photons) should be maximally perturbed by the measurement in order to measure the photon number accurately.

Each individual measurement in this process gives only one answer about the number of photons in the coherent state. As the measurement process is started, these answers are uniformly (randomly) distributed among the possible values, but as the measurement progresses and more and more atoms are sent through the light field, the answer converges (collapses) to a particular value. Starting the process all over again with a freshly trapped, but identically prepared, light field, the measurement converges again to a different number. By repeating the measurement process again and again, the distribution of the number of photons in the initial

coherent state can be reconstructed (Fig. 1).

The ability to follow a quantum measurement stage by stage as the wavefunction collapses opens up broad avenues of approach to fundamental questions that were previously accessible only in simulations. The process of measurement might now be manipulated to steer a system to a particular final state using quantum feedback or to explore conditional quantum dynamics — using the outcome of a measurement to determine a system's evolution. The Parisian team is uniquely placed to advance farther into the quantum jungle, a largely unexplored area where new species of Schrödinger's cat rub shoulders with exotic 'quantum butterflies' — flighty states that flutter between values when disturbed by measurement, only to settle down to rest again. ■

Luis A. Orozco is at the Joint Quantum Institute, Department of Physics, University of Maryland, College Park, Maryland 20742, USA. e-mail: lorozco@umd.edu

1. Gleyzes, S. et al. *Nature* **446**, 297–300 (2007).
2. Guerlin, C. et al. *Nature* **448**, 889–893 (2007).
3. Raimond, J. M., Brune, M. & Haroche, S. *Rev. Mod. Phys.* **73**, 565–582 (2001).
4. Braginsky, V. B. & Khalili, F. Y. *Quantum Measurement* (Cambridge Univ. Press, 1995).



50 YEARS AGO

On the Beach by N. Shute —

THE theme of "On the Beach" is the extinction of the human race resulting from an atomic war. Everybody dies. Just that. When the novel opens in Melbourne, 1962, nobody is alive in the northern hemisphere. Movements of the atmosphere are steadily carrying lethal particles southwards. Mr Shute has deployed his remarkable imagination as engineer, naval officer and storyteller... On his immense popular following this book can only inflict a haunting distress: one takes off one's hat to him. Mr Shute has limited his canvas: he has taken only five main characters—five very ordinary people at that; and he has pitched the emotional and dramatic tone invariably low. Until the moment when the radiation sickness comes on, everybody sticks to the tamest of domestic preoccupations; and then quietly takes a suicide pill. The effect is hypnotic and also odd. Mr Shute's world ends not, as the epigraph warns, with a whimper, but with a stoical silence, movingly impressive... The moral in so far as Mr Shute states it explicitly, comes via his characters thus: "Maybe we've been too silly to deserve a world like this"...

From *Nature*, 24 August 1957.

100 YEARS AGO

According to *Engineering*, an Australian record in wireless telegraphy has been achieved by the successful transmission of messages from H.M.S. *Challenger*, one of the Australian squadron at present stationed in Hobson's Bay, to the flagship *Powerful*, which at the time was moored in Farm Cove, Port Jackson. The *Challenger* was in communication with the flagship by means of wireless telegraphy the whole of her voyage. The longest message was one flashed over a distance of 410 miles in a direct line, and this constitutes an Australian record, as previously never more than 240 miles had been achieved by warships on the Australian station. From *Nature*, 22 August 1907.

50 & 100 YEARS AGO

ECONOMICS

Age, health and wealth

Frances Cairncross

Ageing populations raise the spectre of crippling healthcare costs. Extra spending on medical research might bring healthier, happier older people who work (and pay taxes) for longer. Is that a good investment?

D. LEVENE Many of the changes that will dominate the first half of the coming century are uncertain, but one is not: the global population will grow older. Already those aged 65 or over, who have rarely accounted for more than 2–3% of the populations of most countries, make up 15% of the rich world's inhabitants. The proportion is also rising in poor countries, as women in many nations have fewer children, and life expectancy rises, not just in infancy but in late middle-age and even in old age.

Hallmark Cards now sells about 85,000 100th-birthday cards a year, in three different designs. Reaching the century is all very well — as long as it does not mean spending a quarter of a century incapacitated or in a nursing home. So one of the most pressing issues about the rise in life expectancy at older ages is whether more years of life mean more years of disability or more years of good health.

This matters not only for those with greying hair, but also for governments. A consequence of falling fertility rates is that there will be fewer young people relative to the number of old people. One result is an upward creep in the average age of the labour force; another is a fall in the proportion of workers to non-working old folk. The upshot is fewer taxpayers, but also more people who are not contributing to the economy and are likely to draw increasingly on health services. If the books are to balance, people will need to work for longer. But will their health be good enough for them to do so? And if it is, will they be as productive as younger folk, or will an older workforce in industrialized countries lose its competitive edge against industrializing countries that still have youthful employees?

These matters are the starting point for a new paper¹ on labour-force participation, increases in human capital and the implications for spending on research. The lead author is Kenneth Manton, who has revolutionized the conventional wisdom on health in old age. He is a mathematical demographer who ran the National Long Term Care Survey (NLTCs), a survey that began in 1982 and is designed to provide information about the population of chronically disabled elderly people in the United States. As Manton began to publish results^{2,3}, the initial response was one of incredulity: contrary to the prevailing wisdom of gerontologists, who thought that modern medicine would not cure or prevent the chronic diseases of old age, the results showed that chronic disability



If the economic books are to balance, people will need to work for longer.

among over-65s is declining, and at a gradually accelerating pace. The most recent NLTCs shows that the pace of decline has increased from 0.6% in 1982 to 2.2% in 1999–2004.

Will this transformation continue? If it does, the savings in both human misery and government expenditure will be tremendous, as will the benefits to the world economy. Instead of a growing population of disabled elderly in need of expensive long-term care, many countries can look forward to an army of spry, intelligent and hard-working old folk. The role model might be Alan Greenspan, who went on running the US central bank until he was almost 80.

Manton and his colleagues are optimists. Health quality is increasing faster than life span, they argue, and has been on the rise for much of the past century. The value of this improvement to individuals and to the economy can be quantified. There is every reason to believe, they say, that the improvement can continue. Many of the gains in both physical and mental health of the elderly result from innovations in medical care. If this process continues, the economic gains will grow still larger, greatly reducing worries about the financial burden of an ageing society.

Therefore — and this is the nub of the authors' argument — there are strong financial reasons for governments, and especially the US government, to support hefty investment in medical

research. Even if only 30% of those old people with better health choose to continue working, the authors conclude that the economic gains overall would be sufficient to support a doubling of investment in biomedical research over five years, and a quadrupling over a longer period.

This cheerful conclusion raises several questions, of which I'll mention two. The first is whether Manton and colleagues' optimism is justifiable. Some recent research raises doubts, driven mainly by swelling waistlines in the United States. For instance, the initial results of another, and newer, long-term survey, the Health and Retirement Study, sponsored by the US National Institute of Aging, provides a view not only of people aged 65 and over (as in the NLTCs) but of the older middle-aged. Analysing its results, Cutler and colleagues⁴ conclude that current trends in obesity could reverse about one-third of the gains in life expectancy within the next 30 years.

On the other hand, there clearly remains much scope for medical advance. For a review⁵ of potential biomedical advances by the Centers for Medicare & Medicaid Services, panels of technical experts were asked to identify possible breakthroughs that would affect the health and life expectancy of the elderly. They came up with a list of 33 that are likely to become commonplace during the next two decades. Even obesity may turn out to be less of a challenge than it now seems. The proportion of overweight people will not rise for ever; some of us will manage to keep our figures as we age. And the consequences of obesity are becoming more treatable as the years go by.

The second question is whether Manton *et al.* are right to put so much emphasis on the economic benefits to be gained from medical research. One of Manton's collaborators (and a co-author of this paper) is Arthur Ullian, a wealthy and driven advocate of more spending on health research, who gave up a career in property after a serious accident left him in a wheelchair. The paper¹ sounds a bit like a lobbying document for the US Senate appropriations committee, a body that has sometimes taken the view that expensive health interventions merely drive up costs.

Nonetheless, the argument that the returns on investment in medical-care technology and pharmacology have so far easily outstripped their costs is persuasive. In particular, the development of statins for reducing blood cholesterol levels must be one of the most economically beneficial inventions of the past three decades. But medical research is surely worth conducting for the improvements it brings to human happiness, as well as to productivity. And it is not the only way to extend life expectancy. Changing people's behaviour is also enormously important.

The fall in the number of people smoking and the rise in education have been two powerful influences on healthier life expectancy. But the well-educated have enjoyed much larger gains both in life expectancy and in health in old age

than have the less educated — and the latter find it hardest to abandon their cigarettes. If one is talking about cost-effective ways to keep even more old folk sprightly, then the best bet may be to improve the education of the poor and to persuade them to stop smoking. ■

Frances Cairncross is at Exeter College, University of Oxford, Oxford OX1 3DP, UK. e-mail: frances.cairncross@exeter.oxford.ac.uk

1. Manton, K. G., Lowrimore, G. R., Ullian, A. D., Gu, X. & Tolley, H. D. *Proc. Natl Acad. Sci. USA* **104**, 10802–10807 (2007).
2. Manton, K. G., Stallard, E. & Corder, L. *J. Gerontol. B* **50**, S194–S204 (1995).
3. Manton, K. G., Corder, L. & Stallard, E. *Proc. Natl Acad. Sci. USA* **94**, 2593–2598 (1997).
4. Cutler, D. M., Glaeser, E. L. & Rosen, A. B. *Is the US Population Behaving Healthier?* NBER working pap. 13013 (Nat. Bur. Econ. Res., Cambridge, Massachusetts, 2007).
5. Goldman, D. P. *et al. Health Status and Medical Treatment of the Future Elderly: Final Report* (Rand, Santa Monica, California, 2004).

MATERIALS SCIENCE

Stirring stuff

David J. Pine

Take silicon, soak in water, add acid — and stir. This simple new recipe for the self-assembly of complex microstructures belies an involved sequence of hydrophobic, electrostatic and van der Waals interactions.

In folding its proteins and constructing its complex membranes, nature uses self-assembly: bathed in water or another liquid, tiny building-blocks come together by virtue of their shape and interactions. As they report in the journal *Small*, Onoe *et al.*¹ adapt these natural processes for their own designs. They describe a method for assembling parts just 10 micrometres across into complex, three-dimensional objects, and go on to build up chains of interlocking rings. The research is another step towards the ultimate goal of building electrical or optical circuits, or even microscopic machines, from components at micrometre and smaller scales.

The folding of a protein molecule from a long, linear sequence of linked amino acids is one of nature's more spectacular demonstrations of self-assembly. In a watery environment, certain amino acids along the chain can attract each other by various means — van der Waals attraction, hydrogen bonding or hydrophobic interactions. Others might repel each other through their electrical charges or hydrophilic interactions. Geometrical constraints limit which amino acids along the chain can interact with each other, so which amino acid occupies a given position in the chain is crucial to the final topology and shape of the protein. Geometry thus conspires with attractive and repulsive forces to fold the amino-acid chain into the complex shape that gives a protein its particular function.

Simpler examples of self-assembly also abound. Soap molecules, or 'surfactants', are one. These consist of a water-loving (hydrophilic) head connected to a water-hating (hydrophobic) tail. When placed in water, the hydrophobic tails of different surfactant molecules bunch together to form the core of a sphere, with their hydrophilic heads at the surface. This way, both heads and tails get to live in their desired surroundings. While arranging themselves, the tails often corral a particle of dirt, isolating it so that it can be washed away.

Onoe and colleagues¹ exploit the same ideas of self-assembly, except that their building-blocks are not organic molecules, but particles of silicon formed into slightly tapered cylinders 10 μm in diameter at the wider end. This wider base is coated with a layer of hydrophobic molecules, whereas a thin layer of silicon dioxide (SiO_2) forms naturally through oxidation of the remaining surfaces in air. When placed in water with a nearly neutral pH of 6.5, the SiO_2 surfaces become negatively charged and repel each other when two particles come close. The hydrophobic bases of two cylinders, by contrast, can avoid contact with the water by pairing up flush against each other. And this is exactly what happens when the water is stirred to bring particles close to one another: the hydrophobic surfaces pair up to form barrel-shaped 'dimers' (Fig. 1a).

The authors then add a small drop of acid to the water. At the resulting lower pH of 2.0, the

SiO_2 surfaces lose their charge, and the barrel-shaped dimers no longer repel each other. In fact, when close together they experience an attractive van der Waals force, which is strongest between the exposed flat tops of the dimers, which provide the largest area of contact. (The authors also roughen the round sides of the dimers to diminish the attractive force between the already small contact area of the curved surfaces.) So, when the acid solution is gently agitated, the barrel-shaped dimers undergo a further stage of self-assembly into long, thin cylinders.

The next challenge, making a chain of linked rings, takes the process a significant step further. Starting this time with particles shaped like a wide squat U, with a square hydrophobic patch on the inside of the base, Onoe *et al.* follow the same procedure as before (Fig. 1b). The particles first link up at the hydrophobic patches, and then, when the pH is lowered, connect at their ends by van der Waals attraction to form rings. Only a couple of links are demonstrated, but the ability to form such topologically complex objects is intriguing.

This work follows on from the pioneering experiments of George Whitesides and his colleagues, who exploited hydrophobic and hydrophilic interactions, as well as capillary action, to build complex structures on scales of 10 μm and larger: rings, tetrahedra and tiny rugby balls²; chiral helices³; and stacked plates and three-dimensional crystals⁴. Here too, the geometry of the building-blocks largely determines the objects that are assembled. Particles can even be shaped to fit together in a lock-and-key manner². More recently, this mechanism has been exploited on length scales of a micrometre to build up an 'alphabet soup' of particle shapes⁵. These smaller particles are batted about naturally by thermal brownian motion, self-assembling with no external intervention, such as stirring, at all.

What is especially interesting about Onoe and

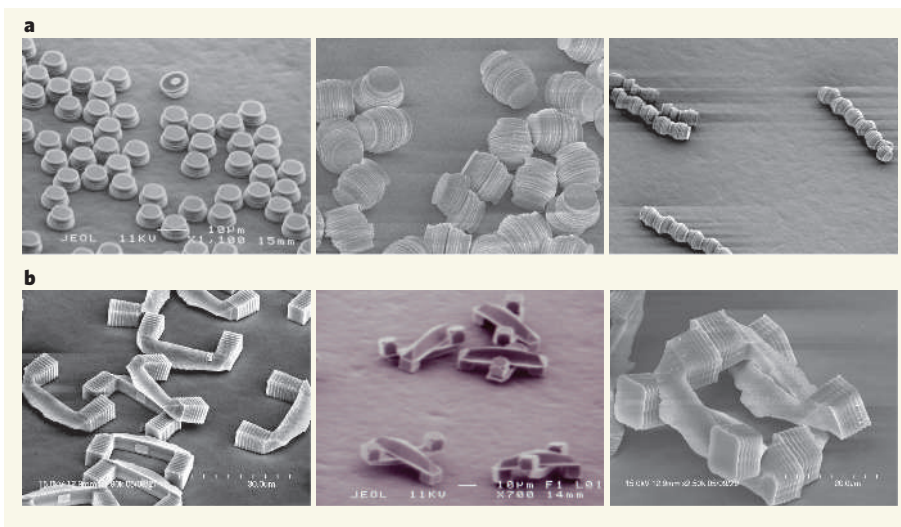


Figure 1 | Self assembly in action. Onoe and colleagues' processes seen under the microscope, starting from (a) tapered cylinders and (b) U-shaped particles. (Images reproduced, with permission, from ref. 1.)

colleagues' technique¹ is the sequential control over a self-assembly process that involves different interactions — first hydrophobic forces, then electrostatic repulsion and finally van der Waals attraction. Sequential self-assembly has been carried out before, most notably in beautiful experiments by Ned Seeman⁶ and Chad Mirkin⁷ and their colleagues using artificial sequences of DNA. But those experiments use the coding of DNA base pairs rather than different forces to control the sequence of self-assembly. Moreover, they are 'biokleptic' (Seeman's term): that is, they borrow heavily from biological processes. Onoe and colleagues' approach is more general and, at the current stage of development, much less powerful. It does, however, represent a useful addition to our toolbox.

To develop self-assembly into a practical technology, we will need to be even more ingenious than nature, exploiting all the interactions

at our disposal and creating yet-to-be envisaged pathways. The recent work¹ takes us a step further down that road.

David J. Pine is at the Center for Soft Matter Research, Department of Physics, New York University, 4 Washington Place, New York, New York 10003, USA.
e-mail: pine@nyu.edu

1. Onoe, H., Matsumoto, K. & Shimoyama, I. *Small* **3**, 1383–1389 (2007).
2. Terfort, A., Bowden, N. & Whitesides, G. M. *Nature* **386**, 162–164 (1997).
3. Breen, T. L., Tien, J., Oliver, S. R. J., Hadzic, T. & Whitesides, G. M. *Science* **284**, 948–951 (1999).
4. Clark, T. D., Tien, J., Duffy, D. C., Paul, K. E. & Whitesides, G. M. *J. Am. Chem. Soc.* **123**, 7677–7682 (2001).
5. Hernandez, C. J. & Mason, C. J. *J. Phys. Chem. C* **111**, 4477–4480 (2007).
6. Chen, J.-H., Kallenbach, N. R. & Seeman, N. C. *J. Am. Chem. Soc.* **111**, 6402–6407 (1989).
7. Taton, T. A., Mucic, R. C., Mirkin, C. A. & Letsinger, R. L. *J. Am. Chem. Soc.* **122**, 6305–6306 (2000).

22 phylogenies with more than 70 species were considered, this percentage increased to 55%. A correlation between phylogenetic relatedness and ecological similarity — the fraction of common interactors between two species — was detectable in 43% (50 of 103) of phylogenies.

For their ten largest communities (those with more than 40 animal and plant mutualists), Rezende *et al.* simulated coextinction cascades by removing pollinators systematically, starting with the least-linked (most specialized) partners and moving to the most-linked (most generalized). Species left without any local interaction were assumed to become extinct. The simulations show that related species do tend to become extinct together, although the effects were small.

Overall, given the modest percentage correlations cited above, Rezende and colleagues find that phylogenetic relationships do not have a marked effect on the degree and strength of interactions in local communities. This result will disappoint those who expect to find such a signal in every nook and cranny. One explanation may lie in the hugely varied temporal scales over which the hundreds of plant and animal species in the 59 local networks have been interacting. The networks encompass Mediterranean, tropical, temperate, subtropical and Arctic communities, which differ greatly in their stability and numbers of evolutionarily old or recent species. Also, some of the communities have comparatively few closely related species, and so lack a strong phylogenetic structure. Given the different temporal and spatial scales over which different pollinators (birds, bees, flies, beetles) and plants (tropical or temperate, woody or herbaceous) evolve, very large networks may be needed to discern phylogenetic signals in interaction strength and degree¹⁰.

However, even then, phylogeny might not predict numbers and kinds of mutualistic interactions. As has been shown¹¹ for a network involving many insects and one focal plant, broad-leaved lavender (*Lavandula latifolia*), having few or many visitors to a flower may not be a trait that is invariant at the species level, but instead may depend on research design (sampling effort) and biological phenomena (variation in absolute and relative pollinator abundance or visitation rates). If a large proportion of the interactions counted in pollination networks are not species-level traits, this would explain the absence of clear phylogenetic effects in insect–flower networks.

There is a risk that treating mutualistic networks as “coevolved structures rather than as diffuse multispecific interactions”⁵ could lead research on networks into a trap from which community ecology has long escaped¹². Instead of revealing coevolved interactions, Rezende and colleagues' results might be taken as showing that such interactions are not very important. And in terms of extinctions, the formation of associations between migrating or invading species and local generalist species

EVOLUTIONARY BIOLOGY

Structure in mutualistic networks

Susanne S. Renner

Statistical analyses of the networks formed by plant–animal mutualisms can now take account of the relatedness of the players on either side. How helpful is this innovation for understanding network dynamics?

The mutually beneficial relationships between plants and animals take several forms. One example is pollination. Another is the process by which a fruit-eating creature, a frugivore, gets a meal and subsequently disperses a plant's seeds in its droppings.

In the context of a local ecological community, such relationships can be seen as networks in which the evolutionary dynamics of the partners may be mutually dependent, contributing to an array of coevolutionary processes¹.

On page 925 of this issue, Rezende *et al.*² report an analysis of plant–pollinator and plant–frugivore networks that includes information on the evolutionary history of the partners — that is, on the phylogenetic relatedness of each partner to other plants or animals in the network. Their aim was to determine whether relatedness affects network structure and whether it predicts ‘cascades’ of coextinction. This is the first such evolutionary network analysis, and it highlights both the power and the limitations of the approach. Two central concepts are those of species ‘strength’ and species ‘degree’ (as described in Box 1, overleaf, which gives the background to the method). Combining the two concepts has produced a boom in the analysis of plant–animal networks^{2–4}.

In mutualistic networks, optional interactions occurring among many species are common; most interactions are strongly asymmetric; and species interact with nested subsets of partners^{1,5,6}. Thus, a large number of species having few inter-

actions coexists with a relatively small number of super-generalists. Relatively rare plants and animals, as well as those with comparatively few partners, interact primarily with a core group of abundant generalist species.

Related participants in mutualisms are likely to have similar morphology, physiology and behaviour. These are traits that evolve as lineages diversify, so we would expect the structure of mutualistic networks to be influenced by the hierarchical phylogenetic relationships present in a particular community. This is the assumption Rezende *et al.*² set out to test. Statistical methods to estimate the role of phylogenies in explaining patterns of trophic (feeding) association are now available^{2,7}; these methods are based on established statistics for the phylogenetic comparative method^{8,9}. The basic approach is to structure the problem of pollinator–flower associations as a statistical model in which phylogenies are used to give the covariance structure of the ‘error’ terms.

Rezende *et al.* applied these methods to 59 plant–pollinator and plant–frugivore networks, which were compiled from the literature. Their approach involved obtaining phylogenies for the insects in 35 networks, birds (all frugivores) in 18, and flowering plants in 52. The number of interactions per species was significantly phylogenetically conservative in 25% of the phylogenies (26 of 105) and a third of the networks. Small phylogenies provide little power to detect phylogenetic effects, and when only the

Box 1 | Methods for quantifying mutualisms

Quantitative approaches to understanding plant-animal mutualisms go back to the application of food-web theory to a large sample of pollination and seed-dispersal networks¹³. Mutualistic interactions can be arranged as a matrix of dependences between $a = 1, 2, \dots, i$ species of animal and $p = 1, 2, \dots, j$ species of plant, with the total number of species in the system, M , equal to $a + p$.

One can then examine the proportion of all possible interactions that actually occur and, for interacting pairs of species, estimate the

relative magnitude of the interaction, or the number of interactions per species (referred to as species 'degree'). The realized number of interactions (connectance) scales in proportion to M ; the potential number of interactions scales as $a \times p$; the average number of interactions per species varies independently of M (ref. 13). Because connectance covaries with species richness (as in other food webs), network studies control for M by doing regression analysis on residuals.

Recently, two further

parameters have been introduced⁵, quantifying mutual dependence or species 'strength'. The first is the dependence of plant species j on animal species i — that is, the fraction of all visits from a particular animal species. The second is the dependence of animal species i on plant species j (that is, the fraction of all visits by this animal species going to a particular plant species). This allows an index of asymmetry to be calculated for each pairwise interaction, paving the way for the latest analyses²⁻⁴.

S. S. R.

these surface-grown films can coat a broad range of substrates, from large flat surfaces to highly curved colloidal particles. Flexible polymers can be layered onto complex, curved geometries — for example, red blood cells, which have a concave structure, have been used as templates for LbL assembly³.

Kreft *et al.*¹ used microspheres of ionic calcium carbonate as the template for the inner capsule of their structures. These spheres offer two advantages as templates for LbL assembly. First, they can be dissolved at the end of the process by washing with a solution of EDTA (a molecule that traps calcium ions); this is a milder way of removing templates than previously reported methods, which involved harsh organic solvents. Second, calcium carbonate can be used as a matrix to trap biomolecules, including proteins and DNA⁴, as the processes involved are mild and 'biofriendly'.

The authors began by forming microspheres of calcium carbonate, using them as a matrix to immobilize protein molecules (Fig. 1, overleaf). They also trapped magnetic nanoparticles in the spheres to provide an easy method of separating intermediates from unwanted co-products during the assembly process. The authors coated the spheres with ten alternating layers of synthetic polymers, and then grew a thick shell of calcium carbonate on the surface of the resulting polymer film; a second kind of protein was incorporated into this layer. The authors deposited more layers of polymers onto the outer shell, and finally removed the calcium carbonate by washing with EDTA. This yielded micrometre-sized, shell-within-shell microcapsules, filled with proteins; like a cell, the proteins in the synthetic 'nucleus' were different from those in the 'cytoplasm'.

Both cell and nuclear membranes are semi-permeable — they permit some molecules to pass, but not others, which allows many essential reactions to be confined to either the nucleus or the cytoplasm. Although small molecules diffuse through synthetic membranes, many proteins and nanoparticles are retained within LbL capsules, so LbL shells can also selectively control which molecules pass through them.

Kreft and colleagues¹ used this effect to compartmentalize an enzyme reaction. They prepared a shell-within-a-shell system in which enzymes in the outer chamber oxidize water molecules to form hydrogen peroxide. The peroxide readily diffuses into the inner chamber, where it oxidizes an encapsulated dye; this produces a red colour that slowly diffuses back into the outer compartment. This proof-of-principle reaction will no doubt spur interest in other embedded reactions, and perhaps stimulate new designs for synthetic cells.

Compartmentalized structures that mimic cells have been generated using other mild processes. One interesting approach is to trap a solution of two polymers within a lipid vesicle⁵; if the polymers and conditions are carefully chosen, a central pool forms in which only one of the polymers is dissolved; this is

might actually encourage the persistence of local species in the long term^{2,5,10}.

Nonetheless, with habitat fragmentation and climate change now occurring so rapidly, mutualistic networks are likely to be severely affected in many places. Setting networks in an evolutionary context might help to predict their level of resilience. Using the approach developed by Rezende *et al.*, future work on networks might incorporate indirect antagonistic interactions affecting mutualistic networks — such as those in which herbivores consume flowers or fruits to the detriment of the plants concerned. ■

Susanne S. Renner is in the Department of Biology, Ludwig-Maximilians-Universität, Menzingerstrasse 67, 80638 München, Germany. e-mail: renner@umsl.edu

- Jordano, P., Bascompte, J. & Olesen, J. M. *Ecol. Lett.* **6**, 69–81 (2003).
- Rezende, E. L., Lavabre, J. E., Guimarães, P. R. Jr, Jordano, P. & Bascompte, J. *Nature* **448**, 925–928 (2007).
- Memmott, J., Craze, P. G., Waser, N. M. & Price, M. V. *Ecol. Lett.* **10**, 710–717 (2007).
- Santamaría, L. & Rodríguez-Girone, M. A. *PLoS Biol* **5**, e31. doi:10.1371/journal.pbio.0050031 (2007).
- Bascompte, J., Jordano, P. & Olesen, J. M. *Science* **312**, 431–433 (2006).
- Bascompte, J., Jordano, P., Melián, C. J. & Olesen, J. M. *Proc. Natl Acad. Sci. USA* **100**, 9383–9387 (2003).
- Ives, A. R. & Godfray, H. C. J. *Am. Nat.* **168**, E1–E14 (2006).
- Felsenstein, J. *Am. Nat.* **125**, 1–15 (1985).
- Felsenstein, J. *Inferring Phylogenies* (Sinauer, Sunderland, Massachusetts, 2004).
- Memmott, J., Waser, N. M. & Price, M. V. *Proc. R. Soc. Lond. B* **271**, 2605–2611 (2004).
- Herrera, C. M. *Am. J. Bot.* **92**, 13–20 (2005).
- Ricklefs, R. E. *Ecology* **87**, S3–S13 (2006).
- Jordano, P. *Am. Nat.* **129**, 657–677 (1987).

MATERIALS SCIENCE

Embedded shells decalcified

Catherine Picart and Dennis E. Discher

Synthetic microcapsules with membrane-bound inner chambers in which chemical reactions can be isolated and controlled have been assembled, layer by layer. Could artificial cells be on the horizon?

The next time you eat an apple, take a moment to consider its hierarchical structure. The thin, durable skin wraps around the delectable fruit, in which can be found seeds; those seeds, in turn, have a coating of their own to encase their contents. Animal cells possess a similar arrangement of embedded compartments: cell membranes confine the machinery of life, including the nuclei with their membrane-ensconced genomes. Both cell and nuclear membranes self-assemble from lipid molecules, and are fortified with networks of proteins whose structures are templated by interactions with the membranes. The hierarchical assembly of a synthetic, cell-like struc-

ture has now been described in *Angewandte Chemie* by Kreft *et al.*¹ They have used microspheres of calcium carbonate as templates to construct polymeric, shell-within-shell capsules. Once the structures are assembled, the templates are simply washed away.

The authors adopted a strategy known as layer-by-layer (LbL) assembly, which entails dip-coating a template alternately in two polymer solutions, one of which is positively charged and the other negatively charged. A wide range of polymers can be deposited in this way, to make multilayered structures with an array of physical and chemical properties, and many potential applications². In addition,

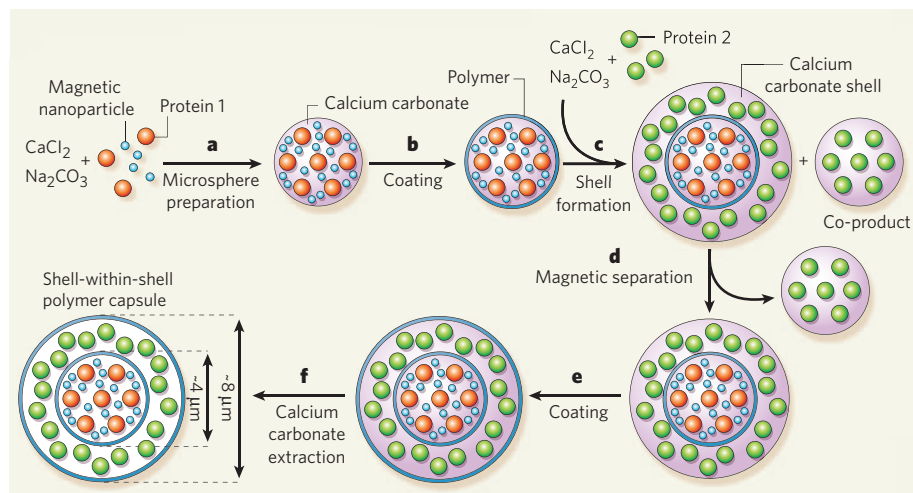


Figure 1 | Polymer shells within shells. Kreft *et al.*¹ have prepared polymeric shell-within-shell capsules. **a**, The authors prepared calcium-carbonate microspheres in which magnetic nanoparticles and a protein were embedded. **b**, The spheres acted as templates for a polymer film. **c**, The authors then deposited a thick shell of calcium carbonate in which a different protein was embedded, and **(d)** separated the desired material from co-products using a magnet. **e**, A second polymer layer was deposited on the outer sphere. **f**, Finally, the authors extracted the calcium carbonate to leave behind a polymeric shell-within-shell structure, with different proteins trapped in separate compartments. (Figure adapted with permission from ref. 1.)

surrounded by a separate solution of the other polymer. Such a system lacks a semi-permeable membrane to separate the two solutions, but has the advantage that the whole thing can easily be triggered to break down using various stimuli, such as heating; this is potentially useful for applications in which the capsule is being used as a delivery system. In comparison, Kreft and colleagues' approach yields physically robust shells, but these could, in principle, be made with degradable polymers.

Kreft *et al.* comment¹ that one of the long-term goals of their work is to produce drug-delivery systems that can release drugs in a

controlled manner, or in response to a biological trigger. For this application, biodegradable multilayered materials are especially useful. Interestingly, several such polymeric materials are about as compliant⁶ as cell and nuclear membranes⁷ — that is, their tendency to revert back to their original shape upon the removal of a distorting force matches that of their biological counterparts. Physical properties such as compliance are surprisingly important in determining the compatibility of materials with biological systems. For example, it is now known that flexible objects are less prone to being engulfed by phagocytic cells, which clear foreign particles

from the bloodstream and often prevent drug-delivery systems from working⁸. The ability to eliminate rigidity from drug-delivery systems makes Kreft and colleagues' decalcification approach¹ even more attractive.

A grander challenge for LbL assembly is to prepare functional mimics of nucleated cells. Perhaps the ingredients for transcribing DNA into messenger RNA could be trapped in the synthetic nucleus of a cell mimic; the RNA could then be translated into proteins in the surrounding 'cytoplasm'. This is undoubtedly ambitious, but such systems might help answer a fundamental question of biology: why do cells store DNA in a nucleus? Regardless of the directions pursued, Kreft and colleagues' mild method for the LbL assembly of nucleated systems and compartmentalized reactions is likely to stimulate the work of materials scientists for years to come.

Catherine Picart is at the Université Montpellier 2, Montpellier, CEDEX 5, 34095, France. Dennis E. Discher is in the Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

e-mail: discher@seas.upenn.edu

1. Kreft, O., Prevot, M., Mohwald, H. & Sukhorukov, G. B. *Angew. Chem. Int. Edn* **46**, 5605–5608 (2007).
2. Tang, Z., Wang, Y., Podsiadlo, P. & Kotov, N. A. *Adv. Mater.* **18**, 3203–3224 (2006).
3. Neu, B. *et al.* *J. Microencapsulation* **18**, 385–395 (2001).
4. Johnston, A. P. & Caruso, F. *Angew. Chem. Int. Edn* **46**, 2677–2680 (2007).
5. Long, M. S., Jones, C. D., Helfrich, M. R., Mangelny-Slavin, L. K. & Keating, C. D. *Proc. Natl Acad. Sci. USA* **102**, 5920–5925 (2005).
6. Richert, L., Engler, A. J., Discher, D. E. & Picart, C. *Biomacromol.* **5**, 1908–1916 (2004).
7. Dahl, K. N., Kahn, S. M., Wilson, K. L. & Discher, D. E. *J. Cell Sci.* **117**, 4779–4786 (2004).
8. Benigno, K. A. & Wang, Y. L. *J. Cell Sci.* **115**, 849–856 (2002).

EARTH SCIENCE

Old diamonds and the upper crust

Ian S. Williams

Was the early Earth a blackened landscape of congealed lava, or was it cool enough for oceans to form? The discovery of diamonds in the oldest-known relics of surface rocks adds new élan to this debate.

The earliest era of Earth's history — between the creation of the planet about 4.5 billion years ago and the formation of the oldest known rocks 500 million years later — is the geological equivalent of the Dark Ages. The formal name given to the period, the Hadean, matches the long-held view that during this time Earth's entire surface, or at least large tracts of it, was a hellish, seething mass of molten lava. In contrast to that picture, recent studies of the oldest known relics of surface rocks, tiny and rare crystals of the trace mineral zircon (ZrSiO₄), indicate that it took less than 200 million years

for Earth's surface to cool sufficiently for water to condense. Menneken and colleagues' discovery¹ (page 917 of this issue) that some of those ancient zircon crystals contain tiny diamonds turns the field on its head once more. The results lend succour to the view that the initial hot phase was long-lived, and will invigorate questioning of the cool-Earth concept.

As Earth grew in the dust cloud around the proto-Sun, probably colliding with a planet the size of Mars during the process, surface temperatures would have reached in excess of 6,000 °C (ref. 2). When the surface eventually

cooled, solid rocks started to form, but what those rocks were like, and when the formation process started, are controversial questions. The problem is one of preservation. Intense meteorite bombardment destroyed much of the early crustal rock, and with it evidence of Earth's early environment. In addition, Earth is an active planet on which crustal rocks are continuously destroyed by weathering and erosion, and recycled deep into the interior through plate tectonics. These processes have been going on for at least 3.8 billion years, and so very few crustal rocks older than 3.7 billion years — and none older than 4.03 billion years — are known.

The only surviving relics of Earth's crust from those first 500 million years of the planet's history are zircon crystals, mostly recovered from slightly less ancient sedimentary rocks³. Zircon is a tough mineral: it is both chemically and physically inert under a wide range of geological conditions. As a result, zircon grains formed in old igneous and metamorphic rocks are commonly recycled unchanged into younger rocks.



Figure 1 | Treasure trove. Menneken and colleagues¹ found diamonds in zircons, the earliest known remnants of Earth's crustal rocks that occur in conglomerates at Jack Hills in Western Australia.

Hadean zircon grains are small (less than 0.3 mm in diameter) and rare: so far only two rocks, both from Western Australia, have been found with an 'abundance' of the grains (which means about 1 part per million by weight). Fortunately for those studying the early Earth, these tiny grains carry a remarkable amount of information about the rocks in which they crystallized. That information can be extracted through a range of modern techniques that allow growth zones in individual zircon grains to be analysed on a scale of a few micrometres.

For example, uranium, thorium and lead isotopes provide an accurate measure of a grain's age³; zoning textures⁴ and patterns of rare-earth elements⁵ distinguish igneous from metamorphic growth; and microinclusions of other minerals show the mineralogy of the original host rock⁵. The more titanium a zircon grain contains, the higher its crystallization temperature was⁶; and the greater the grain's enrichment in the oxygen isotope ¹⁸O, the more the magma from which the zircon crystallized interacted with liquid surface water^{7,8}. The hafnium isotopic ratio ¹⁷⁶Hf/¹⁷⁷Hf indicates whether that magma was purely mantle-derived (high initial value) or contained a crustal component (low initial value)^{8,9}. Finally, xenon isotopes, among them the fission products of the extinct plutonium isotope ²⁴⁴Pu, provide some of the most direct evidence available that Earth formed with a plutonium–uranium ratio similar to that of the Sun¹⁰.

Some of the conclusions reached on this basis have been challenged, principally on the grounds that some of the compositions measured do not accurately reflect the temperature or the primary chemical and isotopic compositions of the magmas from which the zircons crystallized. But a majority view prevails. Hadean zircons are mostly of igneous origin, and crystallized in a series of events that occurred mainly between 4.25 and 4.00 billion years ago; crystallization temperatures averaged only about 680 °C (ref. 8); the host magmas had compositions similar to granites, and were produced by the melting of mixtures of crustal and mantle rocks; and the crustal rocks had interacted with liquid water at or near Earth's surface. All in all,

this is convincing evidence that there were at least some periods during Earth's early history when its surface was cool.

The discovery by Menneken *et al.*¹ of microdiamond inclusions in about 1 in 20 zircons, ranging in age from 4.3 to 3.1 billion years, that were randomly selected from the original Jack Hills conglomerate site in Western Australia³ (Fig. 1), will necessitate a careful reassessment of these conclusions. Diamond is a mineral that forms only at high pressures, whether these are transient (during, for example, a meteorite impact) or sustained (through deep burial, for instance). It is therefore not found in low-temperature granite magmas such as those in which the Hadean zircons are inferred to have crystallized.

Assuming that the zircon did crystallize at temperatures of about 680 °C, pressures exceeding 3.5 gigapascals would have been needed for diamond to be stable in the magma source. This pressure is equivalent to burial more than 100 kilometres below Earth's surface. But the Jack Hills zircons show no other evidence of such high pressures — whether transient in the form of shock features, or sustained in the form of relics of other high-pressure minerals, metamorphic overgrowths or rare-earth element compositions indicative of high-pressure growth. If the evidence of the diamonds is correct and the Hadean zircons did crystallize from magmas at high pressure, then those magmas could not have been crustal melts. This would undermine other inferences based on the assumption that the zircons preserve their original chemical and isotopic compositions — including the deduction that the early Earth was cool.

Menneken *et al.*¹ present two possible explanations for the presence of diamonds in zircons of such a wide age range. Either the diamonds formed during a single high-pressure event about 4.3 billion years ago and were then recycled into younger rocks, or they were incorporated into the Hadean zircons by an undefined process that was repeated several times during the first billion or so years of Earth's history. Neither of these explanations is entirely convincing, and the implications for surface temperatures on

the early Earth are not discussed.

A possible explanation that does not contradict the evidence for a cool early Earth is that the diamond was introduced to the zircon not as diamond, but as graphite precipitated into fractures and other imperfections in the zircon grains. De Corte *et al.*¹¹, for example, have studied microdiamond inclusions in zircon from rocks subjected to ultra-high pressure in the Kokchetav Massif in northern Kazakhstan. They concluded that the original carbon that the diamonds were formed from was precipitated through the reaction of carbon dioxide with a reduced carbon phase such as methane.

One can conceive of a process in which a segment of early crust containing zircon from low-pressure igneous rocks of different ages became impregnated with carbon-rich fluids, possibly produced by reactions involving an atmosphere enriched in carbon monoxide, carbon dioxide and methane¹². Rapid deep burial of those rocks by subduction (horizontal tectonics), sagduction (vertical tectonics) or mantle overturn¹² could then have formed diamond in zircon of various ages during a single event, possibly without changing the zircons' chemical and isotopic compositions. During subsequent uplift, metastable diamond trapped in zircon grains could have been preserved, with some partly reverting to graphite — as indeed is observed¹.

Whether the diamond was indeed produced by such a relatively rapid process, or was of long-lived mantle origin, could be tested by examining its carbon isotopic composition and whether nitrogen is present as single nitrogen atoms (indicating little time spent in the mantle at relatively low temperatures) or nitrogen pairs and tetrahedra (long residence time at relatively high temperatures)¹³. Such tests will be needed to further the lively debate on whether the early Earth was hot, or not. ■

Ian S. Williams is in the Research School of Earth Sciences, The Australian National University, Canberra, ACT, 0200, Australia.
e-mail: ian.williams@anu.edu.au

1. Menneken, M., Nemchin, A. A., Geisler, T., Pidgeon, R. T. & Wilde, S. A. *Nature* **448**, 917–920 (2007).
2. Benz, W. & Cameron, A. G. W. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 61–67 (Oxford Univ. Press, New York, 1990).
3. Compston, W. & Pidgeon, R. T. *Nature* **321**, 766–769 (1986).
4. Cavosie, A. J., Wilde, S. A., Liu, D., Weiblen, P. W. & Valley, J. W. *Precamb. Res.* **135**, 251–279 (2004).
5. Maas, R., Kinny, P. D., Williams, I. S., Froude, D. O. & Compston, W. *Geochim. Cosmochim. Acta* **56**, 1281–1300 (1992).
6. Watson, E. B. & Harrison, T. M. *Science* **308**, 841–844 (2005).
7. Wilde, S. A., Valley, J. W., Peck, W. H. & Graham, C. M. *Nature* **409**, 175–178 (2001).
8. Trail, D. *et al.* *Geochim. Geophys. Geosyst.* **8**, Q06014 doi:10.1029/2006GC001449 (2007).
9. Amelin, Y., Lee, D.-C., Halliday, A. N. & Pidgeon, R. T. *Nature* **399**, 252–255 (1999).
10. Turner, G., Harrison, T. M., Holland, G., Mojzsis, S. J. & Gilmour, J. *Science* **306**, 89–91 (2004).
11. De Corte, K. *et al.* *Island Arc* **9**, 428–438 (2000).
12. Kramers, J. D. J. *J. Geol. Soc.* **164**, 3–17 (2007).
13. Taylor, W. R., Canil, D. & Milledge, H. J. *Geochim. Cosmochim. Acta* **60**, 4725–4733 (1996).

OBITUARY

Daniel Koshland (1920–2007)

Biochemist, institution-builder and editor of *Science*.

In a research career that lasted more than 60 years from 1945 until his death on 23 July, Dan Koshland used his skills as a chemist and highly creative thinker to advance our understanding of proteins and cells. An energetic and caring man, passionate about all aspects of science, his persuasive and insightful leadership greatly strengthened every institution of which he was part.

Koshland decided to pursue science after reading, as a young teenager, Paul de Kruif's collection of biographical essays, *Microbe Hunters*, and *Arrowsmith*, Sinclair Lewis's novel of an idealistic scientist and doctor. After receiving a bachelor's degree in chemistry from the University of California, Berkeley, in 1941, Koshland was recruited to the Manhattan Project, and became a group leader under Glenn Seaborg in the team at the University of Chicago that successfully purified plutonium. When the war ended, he entered graduate school in Chicago under another giant of chemistry, Frank Westheimer, then a young assistant professor. Koshland attributed much of his success to the mentoring and inspiration of these two men.

After a brief stint at Harvard, Koshland moved in 1951 to Brookhaven National Laboratory on Long Island, New York, where he focused on understanding through physical organic chemistry how enzymes catalyse biological reactions. He became fascinated by the puzzle of enzyme specificity: how does an enzyme select only one of many closely related molecules as its substrate?

At the time, the experts believed that Emil Fischer's 'key-lock' or 'template' theory, proposed in the late nineteenth century, had solved the problem. But Koshland realized that this model of a rigid enzyme could not explain why an enzyme catalyst is much less effective when a distant part of the substrate molecule is removed. The correct answer, published by Koshland in 1958 only after numerous rejections, is that enzymes are activated by changing their shape when they bind their substrates, and that process is influenced by the distant part of the substrate. This principle, called 'induced fit', was eventually confirmed fully through detailed structures of enzyme-substrate complexes produced by X-ray crystallography. Induced shape change has proved to be a ubiquitous phenomenon, applying not only to substrate recognition, but also to much of the regulation of enzyme activity in biological systems.

In 1965, Koshland moved back to Berkeley, to the Department of Biochemistry. A year later, he published a classic paper,



'Comparison of experimental binding data and theoretical models in proteins containing subunits', now recognized to provide the most general model for cooperative (and anti-cooperative) responses in proteins. In the 1970s, he and his lab shifted to focus on understanding how cells receive and interpret signals, wisely selecting bacterial chemotaxis — the movement of bacteria in response to changes in the concentrations of certain chemicals — as a system that could be explored in exquisite detail. By rapidly changing the concentrations of the molecules that bacteria detect, Koshland demonstrated that bacterial cells have a 'memory' of their environment. In later years, members of his lab worked out much of the biochemistry of the receptor-protein methylation that underlies this process.

In 1976, Koshland showed that differences between bacterial daughter cells can be caused by stochastic fluctuations inherent in systems driven by small numbers of molecules. How biology deals with such fluctuations is now a focus of research in many laboratories. In 1984, together with Albert Goldbeter, Koshland introduced the concept of 'ultrasensitivity', revealing how large differences can originate from a cascade of enzymatic reactions downstream from an initial signalling event. All this work has changed the way we think about biology; Koshland's mentorship has spawned many leaders in these fields.

While maintaining an active research laboratory, Koshland became chair of the Berkeley biochemistry department in 1973 and began a series of institution-building activities across the university. In the 1980s, he drove a revolutionary reorganization

of Berkeley's 200-strong biology faculty from more than ten different small departments into three large ones. This was a seemingly impossible task to those versed in university politics, but its beneficial effect on the quality of both faculty and student recruitments was immediately apparent. The recent development of biology-focused interdisciplinary science and engineering efforts at Berkeley also owes much to Koshland.

In 1985, Koshland became editor-in-chief of *Science*, a pivotal position for world science that he occupied, as a half-time post, for ten years. He transformed the journal, not only vastly improving the scientific research it published by appointing a cadre of talented full-time editors with PhD degrees, but also greatly increasing the quality and quantity of the sections dealing with science news and science policy. He wrote more than 200 editorials, many in the form of dialogues with 'Dr Noitall' that were testaments to both his originality and his remarkable wit. These duties did not dent his formidable scientific productivity: during his editorship, he authored more than 100 scientific articles stemming from work at his beloved Berkeley laboratory.

Dan Koshland was both a model scientist and a model human being. He exemplified all the values that the idealists among us associate with science. He was completely open and honest, with a generous, happy spirit that made him a pleasure to be around. He derived great pleasure from being helpful to everyone, and he went out of his way to support young scientists. I was amazed to discover late in his life that he, a modest man who always flew in economy class and drove a dented car, was very wealthy, heir to part of the Levi Strauss fortune. In fact, Dan was a great philanthropist for science, helping to finance new buildings at Berkeley and at Haverford College in Pennsylvania, and both designing and endowing a new science museum at the National Academy of Sciences in Washington DC. I had the privilege of working closely with him in designing that museum. It is named after his first wife, Marian (Bunny) Koshland, an outstanding scientist in her own right, to whom he was married for 52 years until her death in 1997.

Dan is survived by five children from that first marriage, and by his second wife, Yvonne, a former college friend whom he married in 2000. Family, friends and colleagues will come together to celebrate the breadth of his contributions to science in a memorial service and symposium at Berkeley on 16 September.

Bruce Alberts

Bruce Alberts is in the Department of Biochemistry and Biophysics, University of California, San Francisco, 600 16th Street, San Francisco, California 94158, USA.
e-mail: balberts@ucsf.edu

Integrating molecular and network biology to decode endocytosis

Eva M. Schmid¹ & Harvey T. McMahon¹

The strength of network biology lies in its ability to derive cell biological information without a priori mechanistic or molecular knowledge. It is shown here how a careful understanding of a given biological pathway can refine an interactome approach. This permits the elucidation of additional design principles and of spatio-temporal dynamics behind pathways, and aids in experimental design and interpretation.

Clathrin-mediated endocytosis (CME) is an important vesicle biogenesis pathway^{1,2}. Cargo is packaged with high fidelity into vesicles of defined size that are surrounded by a coat predominantly made of the protein clathrin and adaptor protein complexes. This coat is visible by electron microscopy and intermediate stages of vesicle formation have been documented (Fig. 1a)³. An extensive protein machinery works alongside the scaffolding clathrin and adaptor protein complexes to build a vesicle (Fig. 1b and Supplementary Tables).

The CME pathway was quickly recognized to be important for nutrient and growth factor entry into cells, and, despite its cargo specificity and regulation, some toxins and viruses can still hijack it to gain entry into cells^{1,4}. At the synapse, clathrin-coated vesicles (CCVs) participate in retrieval of synaptic vesicles following exocytosis, in which potentially multiple copies of more than 10 different transmembrane proteins must be packaged, in appropriate stoichiometries, into vesicles of uniform size to maintain synaptic transmission⁵.

We define CME as a 'pathway' because between the input (cargo) and the output (cargo-containing vesicle) there is an organized set of intermediate sub-reactions. In this analysis when we refer to a pathway we mean a macroscopic series of discrete events with a functional outcome and we do not mean higher order functions like cell division or vesicle trafficking, which are composed of orchestrated interactions of one or more pathways.

The CME interactome

At the morphological level, CME is seen as the pathway of cargo recruitment into patches of membrane, which become progressively invaginated to form CCVs. These vesicles are subsequently detached from the parent membrane and then uncoated (Fig. 1a). On closer inspection, this linear depiction is an oversimplification, because at the molecular level there are many micro-reactions occurring simultaneously in this pathway. For example, cargo recruitment, membrane bending and clathrin-polymerization all happen in parallel in the early stages of vesicle formation⁶.

Appreciating the considerable variety of proteins known to be involved in CME, we set out to learn about the pathway from a network approach. Using biochemical, structural and proteomic data^{6–12}, the major protein interactions for CME can be organized as a pathway protein interactome (Fig. 1b)^{6,10} giving us the 'CME-interactome'. This may not be the complete description of all possible interactors but, given the intensity of work on CME, it is likely that a very large percentage of protein–protein interactions are already described and therefore presented in this interactome.

The basic elements of the interactome (individual proteins and lipids, and so on) are called nodes and their interactions are depicted as links connecting the nodes^{13–15}. Stable protein complexes, like the adaptor protein complex AP2 or a clathrin triskelion, are considered as single nodes (a position that has support in the literature¹⁶). To emphasize the core machinery of endocytosis we peeled away interactions that point outwards from the pathway to distinct processes like exocytosis and actin polymerization (Fig. 1b, grey-shaded areas). Thereby we distinguish our pathway from other network modules¹⁷ that can be connected pathways (for example, exocytosis) or modular attachments ('plug-ins'). Such attachments can be used in many diverse pathways (for example, actin polymerization associated with vesicle scission and motility^{18–20}). Also, for simplicity, in our CME interactome we concentrated on protein nodes. Changes in second messengers like calcium, in phosphorylation levels, and in membrane lipid composition could also be described as nodes in the pathway and changes to these components would have wide implications for this and other trafficking pathways. Disregarding the overall regulation is crucial for identifying the core machinery. These plots deliberately underestimate the ways in which pathways are directly or indirectly sensitive to other pathways in the cell when embedded in a larger network.

The first thing that becomes apparent within this newly generated pathway-network is that clathrin and AP2 have disproportionately more interactions than other proteins in this pathway and so fall into the definition of being 'hubs'^{21,21} (see Fig. 1b).

Hubs have been subdivided into 'party' and 'date' subtypes²². Date hubs are biomolecules that have many partners but interact with them at different times or locations (dynamic hubs), whereas party hubs can interact with many biomolecules at once (static hubs). Date hubs are proposed to organize the proteome, connecting biological processes to each other, whereas party hubs are proposed to function within modules (which may be pathways). Although recognizing the descriptive benefits of 'party hub' and 'date hub' nomenclature we call clathrin and AP2 'pathway hubs' because this does not immediately ascribe any behaviour to these hubs, but simply states that they appear within the studied pathway.

CME pathway hubs

The AP2 hub belongs to the family encoding heterotetrameric adaptor protein complexes. Such adaptor protein complexes have a trunk domain that binds to cargo and lipids, and two appendage domains, positioned on flexible linkers that bind to accessory proteins via two distinct binding sites on each appendage (Fig. 2a)^{7,8,23,24}.

¹MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.

It is these appendage domains that form the protein interaction surface for the many accessory proteins (for binding motifs see Table 1 and Supplementary Tables), when concentrated in forming coated pits¹⁰. Adaptor protein complexes do not self-polymerize and so concentration and stabilization of adaptor protein complexes is via binding partners (as shown schematically for the binding partner Eps15; Fig. 2a).

Clathrin is the second hub in the network (Fig. 1b). Clathrin binds to most of its interaction partners via a β -propeller domain found at the end of each leg of the clathrin triskelion (Fig. 2b). During coated vesicle formation, clathrin becomes concentrated by interactions both with adaptor protein complexes and with accessory proteins^{6,12}.

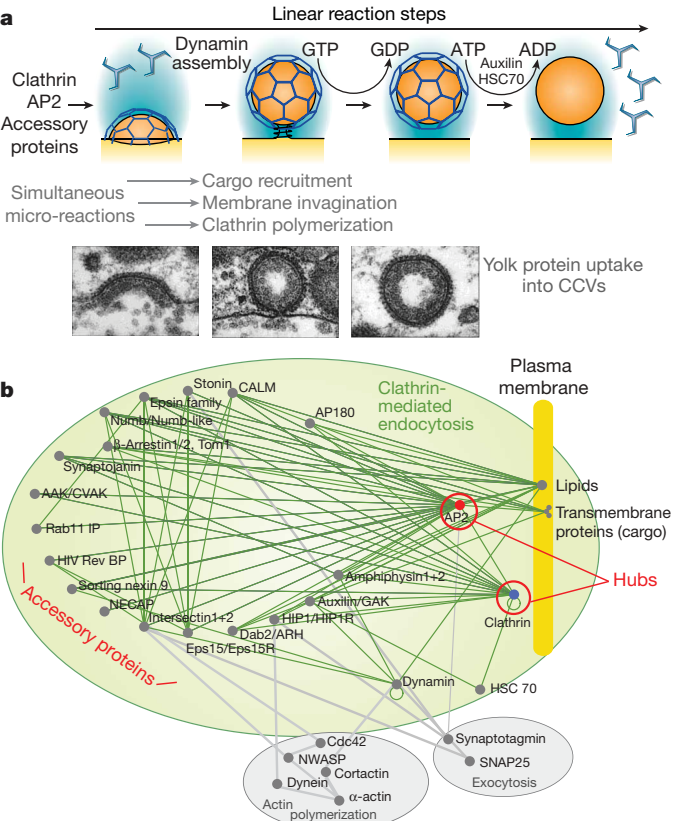


Figure 1 | Clathrin-coated vesicle formation: molecular detail versus a network view. **a**, Budding of a CCV. The initial membrane invagination occurs in concert with cargo recruitment and clathrin polymerization. Accessory proteins, such as AP180/CALM and epsin, are responsible for clathrin recruitment and size-constrained polymerization as well as membrane bending. The AP2 adaptor protein complex links the plasma membrane cargo into the forming clathrin coat. Dynamin is recruited by the accessory proteins amphiphysin, sorting nexin 9 and/or intersectin to the neck of the vesicle to enact membrane scission on GTP hydrolysis, whereas auxilin and the ATPase HSC70 are involved in the uncoating. The stages depicted are also evident in electron microscopy pictures taken from the work on yolk protein endocytosis in chicken³ and reproduced with permission of the Company of Biologists. **b**, The CME interactome: the protein interactions occurring during coated vesicle formation. We have grouped many proteins with similar functions together, for example some of the kinases involved in CME (AAK and CVAK) and alternative cargo adaptors (Dab2 and ARH) are illustrated as single nodes. Different kinds of cargo molecules are also illustrated as one node, which should not lead to the assumption that cargo is a hub. This type of plot shows that AP2 and clathrin are the most common protein interaction points in the network, and these are of importance in the organization of the network, and therefore are surrounded by accessory proteins. The CME interactome is likely to be slightly different in each cell type and will be adapted to the ligands to be endocytosed and the speed with which this is required. Also, differences between species are easy to accommodate within the network, given its flexibility.

Table 1 | Domain structures and functions of CME proteins

Proteins enriched in CME	Proposed function and domain structures	<i>R. norvegicus</i>	<i>D. rerio</i>	<i>D. melanogaster</i>	<i>C. elegans</i>	<i>S. purpuratus</i>	<i>P. falciparum</i>
Clathrin heavy chain	Self-polymerizing support around vesicle	Conservation					
β -Propeller	Clathrin heavy chain repeat	+	+	+	+	+	+
AP2 (α , β 2, μ 2, σ 2)	Links plasma membrane, cargo, clathrin and accessory proteins	Blast searches for α -subunit only					
Trunk	Appendage (α)	+	+	+	+	+	+
Eps15	Scaffolding molecule	+	+	+	+	+	+
Eps15 R	UIMs	+	+	+	+	+	+
Epsin 1	Membrane bending molecule	+	+	+	+	+	+
Epsin 2	ENTH-UIMs	+	+	+	+	+	+
AP180	Membrane binding, clathrin recruitment and vesicle size determination	+	+	+	+	+	+
CALM	ANTH	+	+	+	+	+	+
Amphiphysin 1	Dynamin recruitment	+	+	+	+	+	+
Amphiphysin 2	N-BAR SH3	+	+	+	+	+	+
Dynamin 1	Scission molecules	+	+	+	+	+	+
Dynamin 2	GTPase PH-GED-PRD	+	+	+	+	+	+
Dynamin 3		+	+	+	+	+	+
β -Arrestin 1	Alternative adaptors for GPCR receptors	+	+	+	+	+	+
β -Arrestin 2	Arrestin-ArrestinX	+	+	+	+	+	+
ARH	Alternative adaptor for the LDL receptor	+	+	+	+	+	+
Dab2	Alternative adaptor for the LDL receptor	+	+	+	+	+	+
Numb	Alternative adaptors for the Notch receptor	+	+	+	+	+	+
Numb-like	PTB	+	+	+	+	+	+

Hub proteins are conserved; but where there are multiple isoforms of accessory proteins in mammals and fish, one seems to be sufficient in invertebrates, even in those with nervous systems. A more comprehensive version of this Table is given in Supplementary Tables 1 and 2. Clathrin interaction motifs (X) are incomplete as they are difficult to detect given their sequence variations. Question marks indicate uncertainty. When we do not find clathrin or adaptin interaction motifs in homologues then we generally assume that the protein is not involved in CME. For amphiphysin, we know that a *Drosophila* form of amphiphysin does not have any clathrin or adaptor interaction and does not function in CME and so is not annotated in this Table (see asterisk). ANTH, AP180 N-terminal homology; AP180, adaptor protein 180 kDa; ARH, autosomal recessive hypercholesterolaemia; BAR, Bin/amphiphysin/Rvs; CALM, clathrin assembly lymphoid myeloid leukaemia protein; *C. elegans*, *Caenorhabditis elegans*; Dab2, disabled 2; *D. melanogaster*, *Drosophila melanogaster*; *D. rerio*, *Danio rerio*; EH, Eps15 homology; ENTH, epsin N-terminal homology; Eps15/R, epidermal growth factor receptor pathway substrate 15/related; *P. falciparum*, *Plasmodium falciparum*; PRD, proline rich domain; PTB, phosphotyrosine binding; *R. norvegicus*, *Rattus norvegicus*; SH3, Src homology 3; *S. purpuratus*, *Strongylocentrotus purpuratus*; UIM, ubiquitin interacting motif.

Concentrated clathrin self-polymerizes into a lattice, forming an interaction surface (beneath the outer shell of clathrin legs) made up of many β -propeller domains (Fig. 2b), allowing the lattice to bind to many adaptor protein complexes and accessory proteins simultaneously. This, and clathrin's interaction with itself, provides the required stability to form the budding vesicle and is also likely to be important in the concentration of cargo into the coated pit and in the binding of proteins involved in uncoating.

Functional versus connectivity view of CME

Interactomes give a static picture of pathways that we know to be dynamic²⁵. Thus we attempt to show the changes in connectivity

associated with the different stages of vesicle formation (Fig. 3). In the static interactome (Fig. 1b), we present a network for plasma membrane CCV formation where AP2 is the major heterotetrameric adaptor complex involved in cargo binding²³. We now further focus on one form of plasma membrane CCV formation—that of synaptic vesicle retrieval in nerve terminals²⁶—to emphasize the dynamics of what can happen in one vesicle, but the dynamics of the network will be similar for all types of CCV formation (Fig. 3).

To form a cargo-containing vesicle, cargo needs to be concentrated at patches on the membrane of the appropriate lipid environment. This is likely to happen by the clustering of adaptor protein complexes bound to phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P₂) on the plasma membrane²⁷ (Fig. 3a, functional view), which become stabilized by accessory proteins that bind multiple AP2s simultaneously¹⁰ and also aid in recruitment of clathrin. Clathrin polymerization leads to coat formation. AP2 loses its importance at this stage because clathrin now drives vesicle formation (Fig. 3c, functional view). As soon as the formation of the coated vesicle is completed, it is detached from the parent membrane, and subsequently uncoated (Fig. 3d, functional view).

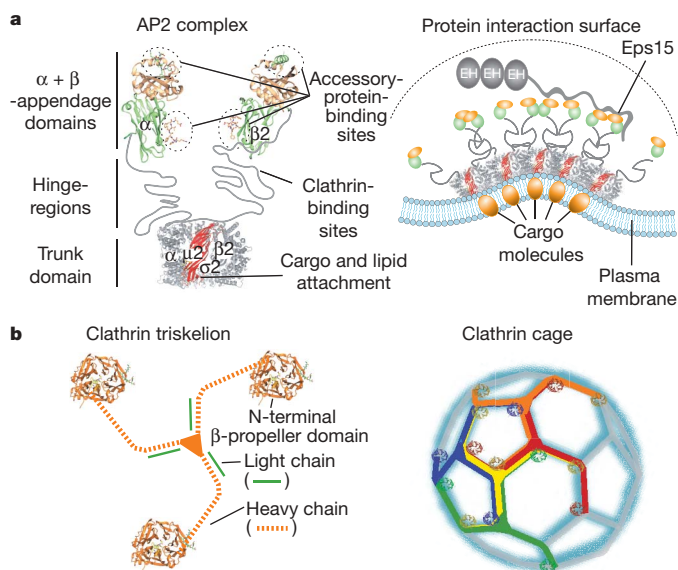


Figure 2 | Structural organization of the hubs in CME. **a**, The major adaptor protein complex in brain-derived CCVs is the AP2 complex. This is composed of four subunits (α , β_2 , μ_2 , σ_2). Membrane cargo interacts predominantly with the μ -subunit (in red), whereas the appendage domains (orange and green), are responsible for binding to accessory proteins (four different peptide interactions are highlighted by dotted lines). These domains are positioned on long flexible linkers and thus they are perfectly designed to be protein recruitment domains bringing the needed partners from the cytosol to the forming coated pit. The right panel shows a representation of a cargo-containing plasma membrane patch accommodating closely packed AP2 complexes, generating a dense appendage protein interaction surface. AP2s are stabilized by clustering molecules such as Eps15, which have been shown to bind up to four appendages simultaneously. **b**, A clathrin triskelion consisting of three clathrin heavy and light chains. Clathrin interacts with ligands via its N-terminal β -propeller domain. When clathrin self-polymerizes into a cage (right panel), these domains form a dense interaction surface composed of at least 180 β -propeller domains per vesicle². From a structural point of view, these propellers are tucked underneath the forming cage (unlike the adaptor protein complex appendage domains that point away from the vesicle) and are thus more likely to have an organizational role in coated pit formation, rather than a protein recruitment role. The clathrin cage on the right is not modelled on a 50 nm vesicle but comes from cryo-electron microscopy reconstruction of isolated cages⁴². The protein data bank accession numbers for the structures are as follows: α -appendage, 1W80; β -appendage, 2IV8, 2IV9; AP2 core, 1GW5; clathrin β -propeller, 1C9L and 1UTC.

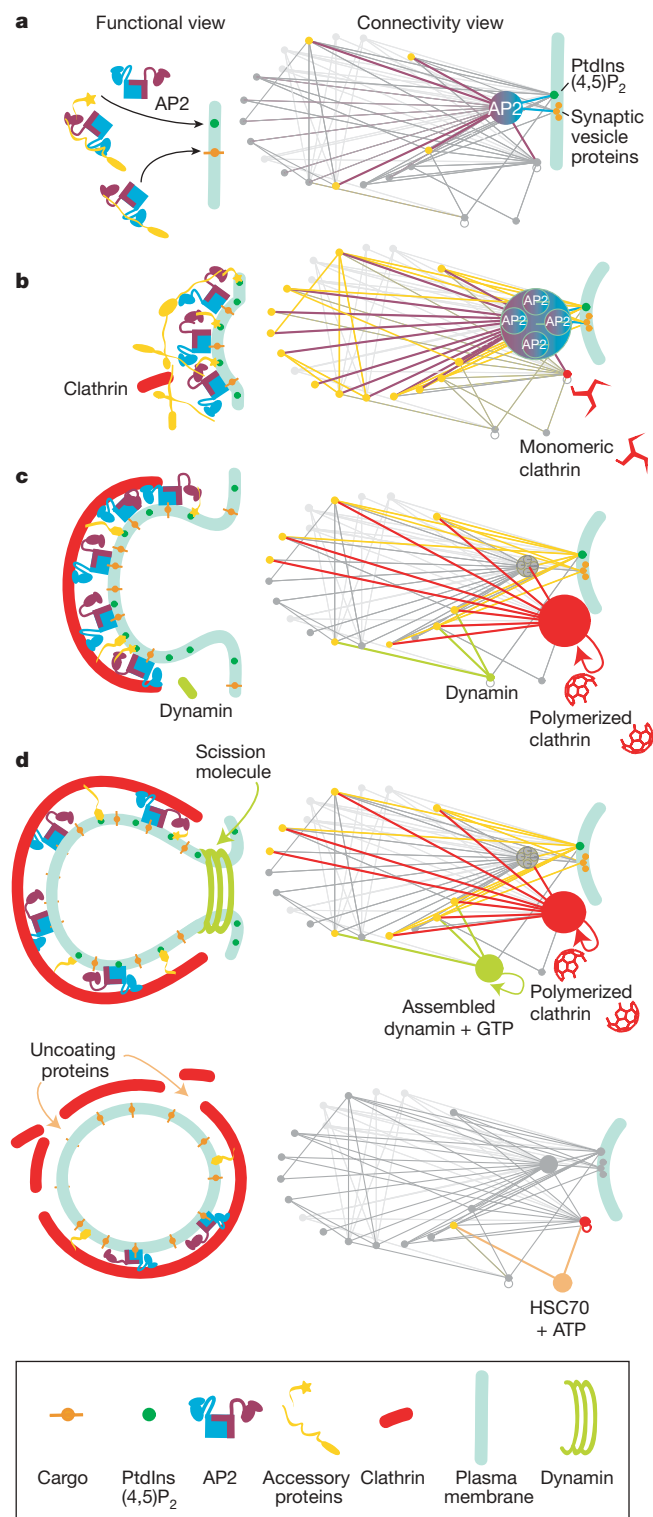


Figure 3 | The steps of synaptic vesicle CCV formation. Functional and connectivity views of CCV formation in nerve terminals. In the network view of Fig. 1b, we included all plasma membrane accessory proteins, but here we are concerned with the changing interactions during synaptic CCV formation (non-synaptic CCV nodes are in light grey). **a**, AP2 recruitment to membranes, by cargo and lipid interactions. **b**, Building the AP2 hub assembly zone: the AP2 pathway hub is stabilized by cross-linking through accessory proteins. **c**, Changing hubs: clathrin is recruited and polymerization stabilizes the forming vesicle, displacing the AP2 recruitment hub and becoming the new organizing hub. **d**, Assembled dynamin and other late-stage interaction partners (like uncoating factors) begin to function.

Connectivity interactomes are a different presentation of pathway progression, emphasizing the dynamic nature of the CME-network (Fig. 3, connectivity view). With each step of CME there is a corresponding change in protein connectivity. The increasing number of purple lines from Fig. 3a to 3b (connectivity view) illustrates the increasing number of interactions AP2 can achieve by being clustered on the membrane. One AP2 molecule can at most interact with four accessory proteins at once, one clathrin molecule, and a limited number of cargoes and lipid head groups. By assembling many AP2s at the membrane these interactions expand and lead to the fully engaged interactome. We would suggest that AP2 only functions as a hub when it is concentrated, forming an AP2 hub assembly zone⁶.

Many of the accessory proteins successfully recruited by AP2 to the site of endocytosis also have interactions with each other and with lipids (yellow lines), further stabilizing this intermediate and allowing the recruitment of proteins that are present in lower concentrations and that bind with lower affinities. In addition, many accessory proteins have clathrin interaction sites and therefore they can participate (with AP2) in clathrin recruitment and concentration, leading to polymerization. Because there is steric overlap between most clathrin and AP2 sites in accessory proteins, AP2 loses its significance and there is a rewiring of the network because the newly formed clathrin hub takes over from the AP2 hub as the organizing centre (red lines to polymerized clathrin; Fig. 3c, connectivity view). Finally the conditions are ripe for the assembled scission molecule dynamin and subsequently the uncoating molecule auxilin to function (Fig. 3d). More information on predicting the timescale of events from topological properties of interactomes can be found in Supplementary Information Part 1 and Supplementary Fig. 2.

What we learn from integration

Plotting the CME pathway protein interactome allows one to immediately identify the key players (hubs) and approximately where they work in the pathway.

Knowing what status a protein has in the network (hub versus a regular node) allows us to predict experimental outcomes. It has previously been suggested that depletion of proteins with many interaction partners (by RNA interference) is more likely to give phenotypes²⁸. Complete depletion of a pathway hub will probably eliminate the pathway. Partial depletion of the AP2 hub affects AP2-specific cargo internalization²⁹. Additional modules (like alternative cargo adaptors) lead to robustness in the network allowing alternative cargoes to be taken up with less or no dependence on AP2. Depletion of regular nodes is likely to be less deleterious than depletion of hubs. Depletion of a protein positioned in-between pathways (like dynamin, see below) may well inhibit many different pathways^{30,31}.

We would suggest that a good test for proteins that function as hubs only when clustered in space and time is that overexpression of such a protein should not have any phenotype on the pathway. Conversely, overexpression of nodes that bind directly to hubs would be predicted to have disastrous effects owing to titration of hub interaction points (see Supplementary Information Part 2, for further discussion).

Fine-tuning hub nomenclature

We describe the pathway hubs, AP2 and clathrin, as the control centres of the CME pathway. We do not ascribe hub status to dynamin in the CME interactome because it has only four links into this pathway. However on the Human Protein Reference Database (<http://www.hprd.org>) dynamin is reported to have approximately 50 interaction partners. Thus, in a proteome network, dynamin would also be a hub and would seem to connect many different pathways together. This is misleading, because from a functional point of view dynamin is simply a module ('plug-in') that is attached to many different pathways in a cell because of its function as a scission molecule, and it is not the connection point between these

pathways. According to the network nomenclature²², dynamin would be called a date hub, but in our view these hubs do not necessarily organize the proteome. On this point we would urge caution because the network nomenclature has preceded a functional understanding of hubs. In a network, the node for dynamin should be present in as many copies as there are pathways in which it is involved, and each of these nodes should not be linked to each other.

We would suggest that hubs could alternatively be named according to where they are positioned in a pathway at a given time, rather than the number and type of interactions. The notion that deletion of party/pathway hubs has much more limited effects than deletion of date hubs²² can now be explained, because deletion of date-hubs will lead to more severe effects on cell viability owing to the additive result of disrupting a frequently used module (for example, dynamin^{30–32}) and thus affecting disparate pathways, whereas deletion of a pathway hub (for example, AP2; refs 29, 33) will have more limited consequences (see also ref. 34 and Supplementary Information Part 2).

Hubs in proteomes are generally not directly connected to other hubs³⁴ and this may mean that the organization of each pathway is independent, with the loss of one hub not affecting other pathways in the cell²². A more recent observation was that party hubs, being the core of highly clustered functional modules, are frequently directly connected²¹. This fits with the direct connection between AP2 and clathrin in the CME interactome. A useful functional feature of this direct connection is seen in the biochemical observation that the clathrin hub displaces the AP2 hub interaction partners and thus ensures directionality to the dynamic protein interactions occurring during CCV formation⁶.

One important property of the pathway hubs AP2 and clathrin is that neither are hubs at the beginning of CME, but mature into hubs by clustering either on the membrane or through polymerization. It is likely that many pathway/party hub proteins will oligomerize or cluster to function as pathway hubs. 'Clustered hubs' are a new subtype of hubs not previously described (Supplementary Information Part 3 and Supplementary Fig. 3).

Consequences of hub-centric pathways

The hub-centric organization of a pathway has at least four major consequences, described below.

(1) Ease of module attachment. An appealing feature of a network is the ease with which modules of functionality can be attached/evolve. Hubs act as the control points where these modules can be plugged into the system to give extra capabilities. In the CME interactome, we find additional alternative cargo adaptors/clathrin-associated sorting proteins (CLASPs) (see Supplementary Tables) added to the network by binding to the AP2 and clathrin hubs, when specific cargo needs to be endocytosed^{6,12,35}. Thus by adding a small module, the repertoire and potentially independent regulation of this specialist cargo³⁶ is accommodated while still using the same overall network. Not all pathways will have such complicated networks. For example, in COPI and COPII vesicle budding the network of interactions is much simpler²³. Simpler and less flexible networks should have hubs with fewer connecting partners and not all hubs can be expected to be regulated by oligomerization. Indeed post-translational modifications such as phosphorylation could well be regulators. Phosphorylation controls the overall flux through CME at the synapse by influencing the ability to form protein networks and so phosphorylation in this context is a master regulator of CME, but it is also clear that phosphorylation can fine-tune individual steps^{27,37,38}.

Modularity can also be seen from looking at the conservation of the CME network across species. We observe that most participating proteins are conserved in the animal kingdom (Table 1 and Supplementary Information Part 4). Species evolutionary distant from mammals, like the malaria parasite *Plasmodium falciparum*, tend to only conserve the hub proteins and not other nodes, an observation previously made more generally from proteome analysis³⁹.

(2) **Dynamic instability.** In a network of interacting proteins most interactions are of necessity low affinity, and from biochemical measurements of affinities we know that many individual steps are dependent on coincidence of multiple interactions^{10,27}. Even the initial event of adaptor protein complexes binding to cargo is dependent on the presentation of binding sequences in membrane cargo and binding of the adaptor protein complexes to a specific lipid. Coincidence is seen at all stages in the process because affinities are frequently low enough that multiple interactions are required to give an avidity that can allow an interaction to occur. This gives rise to a dynamic instability in the network, because the network assembly will depend on whether sufficient adaptor protein complexes are recruited and sufficient accessory proteins are present to crosslink and organize these molecules and, subsequently, whether sufficient clathrin is recruited to stabilize the forming coated pit.

(3) **Pathway progression.** Dynamic instability leads to the question of how directionality can be imposed on the overall pathway of CCV formation. At some point there must be commitment to coated pit assembly and the pathway needs directionality to reach the end product of a free vesicle.

In CME, we are concerned with a network of interactions that can only form when the first hub protein AP2 is recruited and concentrated on the membrane at sites for endocytosis. The strengths of network interactions around clustered AP2s are described as avidities, as they are the combination of many different interactions. To progress in vesicle formation, these avidities have to be overcome. Normally, this would be overcome with a still stronger interaction, or by weakening the previous interaction (perhaps via post-translational modification). In CME, clustered adaptor protein complexes, cross-linked by high avidity interactions with accessory proteins, need to progress to a mature CCV. In immunofluorescence experiments, it is observed that clathrin is recruited and polymerized around nascent vesicles while accessory proteins (for example, eps15; ref. 6) are lost. This is reinforced by the absence of most accessory proteins from purified CCVs⁴⁰. Biochemical experiments on the other hand tell us that clathrin competes poorly with accessory proteins for AP2 (refs 6, 8, 12) and so it would seem that clathrin is unlikely to be responsible for this loss of accessory proteins. The answer may lie in the different polymerization states of clathrin: in the biochemical competition experiments, clathrin is not polymerized but on the vesicle, clathrin is polymerized (Fig. 3b, c). Polymerized clathrin will have different properties and needs to be viewed as a semi-solid matrix. Neither affinity nor avidity interactions describe the interactions of this matrix with its binding partners sufficiently, and therefore we have previously described this mass-action as 'matricity'. Matricity describes a situation in which the matrix does not have traditional diffusion kinetics, and thus a very low-affinity interaction appropriately positioned may well be highly significant. Thus, the network moves from affinity-based interaction, to avidity-based interaction, to matricity-based interaction. This gives directionality to the pathway (Supplementary Fig. 4).

One problem posed by this increasing stability of the CME interactome is that eventually components have to be recycled and reused for the next CCV formation. It is at this point that ATP hydrolysis by the uncoating proteins gives the required injection of energy to re-prime the molecules so that they can start the cycle again, when required. Of course, there are many other required inputs of energy into the system, such as in making the appropriate lipids, but it is interesting that this major input of energy occurs after vesicle-formation is committed (dynamin GTP hydrolysis) or completed (ATP hydrolysis by HSC70 during uncoating) (Fig. 3d). In exocytosis, we also see a network of protein interactions in which ATP hydrolysis is mostly seen after vesicle fusion has occurred, and this is used to re-prime the molecules for the next round of fusion. This use of ATP late in the cycle also adds to the dynamic instability of the pathway.

(4) **Fidelity and flexibility.** In CME, there is an interesting observation that hubs, AP2 and clathrin, bind to overlapping partners, and both hubs compete with each other for these partners. Evolution has neither resulted in these hubs being dispensed with, nor combined them into one, because the AP2 hub also has a specialized function of specific cargo interaction and clathrin has the specific cage-forming function. If these hubs were to become one, then the pathway would lose the dynamic instability, and hence the directionality and flexibility. We argue that AP2 binding modules must surely result in the expansion of cargo recruitment possibilities and if cells had no adaptors, but clathrin were to include a cargo binding domain, then this would reduce the flexibility of the cargo repertoire. Also, if clathrin were to polymerize on cargo binding then the process is more likely to cascade in a forward direction because clathrin self-polymerizes, forming empty cages. Natural selection has, instead, resulted in non-self-polymerizing modules to capture cargo, and a non-membrane-binding module to form the polymerized coat around this module. Thus the pathway has in-built fidelity.

Conclusion

Networks and interactome pathway maps are useful tools to help explain a biological pathway, but they need to be anchored in biological contexts and experimental fact. This can initially be done by embedding interactomes within well-characterized pathways and using this information to inform us in other pathway investigations. We have made many observations about the nature of the CME interactome—about what constitutes a hub, the modularity of the system, the possible evolution of the network, and about dynamic instability, directionality and robustness within networks. We need to understand more about how proteins function together and are organized into pathways, and how these pathways are integrated into yet larger contexts. For this, a fresh perspective and a vision for conceptual thinking, coupled with technical advances to map out dynamic interactomes⁴¹ and in visual immunoprecipitation²⁵, is needed, which ultimately will lead to a better understanding of complex biological processes.

- Kirchhausen, T. Clathrin. *Annu. Rev. Biochem.* **69**, 699–727 (2000).
- Marsh, M. & McMahon, H. T. The structural era of endocytosis. *Science* **285**, 215–220 (1999).
- Perry, M. M. & Gilbert, A. B. Yolk transport in the ovarian follicle of the hen (*Gallus domesticus*): lipoprotein-like particles at the periphery of the oocyte in the rapid growth phase. *J. Cell Sci.* **39**, 257–272 (1979).
- Marsh, M. & Helenius, A. Virus entry: open sesame. *Cell* **124**, 729–740 (2006).
- Brodin, L., Low, P. & Shupliakov, O. Sequential steps in clathrin-mediated synaptic vesicle endocytosis. *Curr. Opin. Neurobiol.* **10**, 312–320 (2000).
- Schmid, E. M. et al. Role of the AP2 β -appendage hub in recruiting partners for CCV assembly. *PLoS Biol.* **4**, e262 (2006).
- Owen, D. J. et al. A structural explanation for the binding of multiple ligands by the α -adaptin appendage domain. *Cell* **97**, 805–815 (1999).
- Owen, D. J., Vallis, Y., Pearse, B. M., McMahon, H. T. & Evans, P. R. The structure and function of the β 2-adaptin appendage domain. *EMBO J.* **19**, 4216–4227 (2000).
- Blondeau, F. et al. Tandem MS analysis of brain CCVs reveals their critical involvement in synaptic vesicle recycling. *Proc. Natl Acad. Sci. USA* **101**, 3833–3838 (2004).
- Praefcke, G. J. et al. Evolving nature of the AP2 α -appendage hub during CCV endocytosis. *EMBO J.* **23**, 4371–4383 (2004).
- Mishra, S. K. et al. Dual engagement regulation of protein interactions with the AP-2 adaptor α appendage. *J. Biol. Chem.* **279**, 46191–46203 (2004).
- Edeling, M. A. et al. Molecular switches involving the AP-2 β 2 appendage regulate endocytic cargo selection and clathrin coat assembly. *Dev. Cell* **10**, 329–342 (2006).
- Barabási, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Rev. Genet.* **5**, 101–113 (2004).
- Alon, U. Biological networks: the tinkerer as an engineer. *Science* **301**, 1866–1867 (2003).
- Albert, R. Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947–4957 (2005).
- Gavin, A. C. et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
- Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402** (suppl.), C47–C52 (1999).

18. Yarar, D., Waterman-Storer, C. M. & Schmid, S. L. A dynamic actin cytoskeleton functions at multiple stages of clathrin-mediated endocytosis. *Mol. Biol. Cell* **16**, 964–975 (2005).
19. Merrifield, C. J., Feldman, M. E., Wan, L. & Almers, W. Imaging actin and dynamin recruitment during invagination of single clathrin-coated pits. *Nature Cell Biol.* **4**, 691–698 (2002).
20. Kaksonen, M., Toret, C. P. & Drubin, D. G. A modular design for the clathrin- and actin-mediated endocytosis machinery. *Cell* **123**, 305–320 (2005).
21. Ekman, D., Light, S., Bjorklund, A. K. & Elofsson, A. What properties characterize the hub proteins of the protein–protein interaction network of *Saccharomyces cerevisiae*? *Genome Biol.* **7**, R45 (2006).
22. Han, J. D. *et al.* Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430**, 88–93 (2004).
23. McMahon, H. T. & Mills, I. G. COP and CCV budding: different pathways, common approaches. *Curr. Opin. Cell Biol.* **16**, 379–391 (2004).
24. Brett, T. J. & Traub, L. M. Molecular structures of coat and coat-associated proteins: function follows form. *Curr. Opin. Cell Biol.* **18**, 395–406 (2006).
25. Niethammer, P. *et al.* Discrete states of a protein interaction network govern interphase and mitotic microtubule dynamics. *PLoS Biol.* **5**, e29 (2007).
26. Slepnev, V. I. & De Camilli, P. Accessory factors in clathrin-dependent synaptic vesicle endocytosis. *Nature Rev. Neurosci.* **1**, 161–172 (2000).
27. Honing, S. *et al.* Phosphatidylinositol-(4,5)-bisphosphate regulates sorting signal recognition by the clathrin-associated adaptor complex AP2. *Mol. Cell* **18**, 519–531 (2005).
28. Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
29. Motley, A., Bright, N. A., Seaman, M. N. & Robinson, M. S. Clathrin-mediated endocytosis in AP-2-depleted cells. *J. Cell Biol.* **162**, 909–918 (2003).
30. Kruchten, A. E. & McNiven, M. A. Dynamin as a mover and pincher during cell migration and invasion. *J. Cell Sci.* **119**, 1683–1690 (2006).
31. Wienke, D. C., Knetsch, M. L., Neuhaus, E. M., Reedy, M. C. & Manstein, D. J. Disruption of a dynamin homologue affects endocytosis, organelle morphology, and cytokinesis in *Dictyostelium discoideum*. *Mol. Biol. Cell* **10**, 225–243 (1999).
32. Praefcke, G. J. & McMahon, H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nature Rev. Mol. Cell Biol.* **5**, 133–147 (2004).
33. Hinrichsen, L., Harborth, J., Andrees, L., Weber, K. & Ungewickell, E. J. Effect of clathrin heavy chain- and α -adaptin-specific small inhibitory RNAs on endocytic accessory proteins and receptor trafficking in HeLa cells. *J. Biol. Chem.* **278**, 45160–45170 (2003).
34. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
35. Shenoy, S. K. & Lefkowitz, R. J. Seven-transmembrane receptor signaling through β -arrestin. *Sci. STKE* **2005**, cm10 (2005).
36. Puthenveedu, M. A. & von Zastrow, M. Cargo regulates clathrin-coated pit dynamics. *Cell* **127**, 113–124 (2006).
37. Marks, B. & McMahon, H. T. Calcium triggers calcineurin-dependent synaptic vesicle recycling in mammalian nerve terminals. *Curr. Biol.* **8**, 740–749 (1998).
38. Slepnev, V. I., Ochoa, G. C., Butler, M. H., Grabs, D. & De Camilli, P. Role of phosphorylation in regulation of the assembly of endocytic coat complexes. *Science* **281**, 821–824 (1998).
39. Prachumwat, A. & Li, W. H. Protein function, connectivity, and duplicability in yeast. *Mol. Biol. Evol.* **23**, 30–39 (2006).
40. Mills, I. G. *et al.* EpsinR: an AP1/clathrin interacting protein involved in vesicle trafficking. *J. Cell Biol.* **160**, 213–222 (2003).
41. Schubert, W. *et al.* Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. *Nature Biotechnol.* **24**, 1270–1278 (2006).
42. Smith, C. J., Grigorieff, N. & Pearse, B. M. Clathrin coats at 21 Å resolution: a cellular assembly designed to recycle multiple membrane receptors. *EMBO J.* **17**, 4943–4953 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. Additional information can be found at <http://www.endocytosis.org>.

Acknowledgements We thank all those who contributed; M. Ford, G. Doherty and R. Mittal for many ideas that go well beyond what is written; P. Evans and M. Ford for their help with the figures; M. Babu for giving us an interest in network biology and informing that interest; and the many who have read and critically commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to H.T.M. (hmm@mrc-lmb.cam.ac.uk).

MAGNETOELECTRICS

Is CdCr_2S_4 a multiferroic relaxor?Arising from: J. Hemberger *et al.* *Nature* **434**, 364–367 (2005)

Materials showing simultaneous ferroelectric and magnetic ordering are attracting a great deal of interest because of their unusual physics and potential applications¹. Hemberger *et al.*² have reported relaxor-like dielectric properties and colossal magnetocapacitance (in excess of 500%) for the cubic spinel compound CdCr_2S_4 and related isomorphs^{3–5}, concluding that CdCr_2S_4 is a multiferroic relaxor. We argue here, however, that their results might also be explained by a conductive artefact.

The polarization hysteresis loops of Hemberger *et al.*² have a shape unlike that seen in ferroelectrics and more like that of a lossy dielectric^{6,7}, such as the semiconducting thiospinels. The ‘remanent’ polarization is much larger than the pyroelectric one, whereas switchable polarization must always be smaller than the pyroelectric polarization — which on its own calls the hysteresis results into question. The pyroelectric polarizations of 0.05–0.07 $\mu\text{C cm}^{-2}$ are themselves tiny for any proper ferroelectric, and are consistent with thermally stimulated discharge of space-charge injected during poling⁶.

Ab initio calculations⁸ preclude soft phonons, and hence proper ferroelectricity. Nor have Raman experiments⁹ found soft phonons or a lack of centrosymmetry. Raman results with wavelengths near or below the resonant edge¹⁰ are consistent with local non-centro-

symmetry (a necessary, although insufficient, condition for ferroelectricity) — however, this technique is unreliable because nominally forbidden odd-parity longitudinal-optical phonon scattering can be strong even in centric crystals¹¹ when there is real absorption.

Broad dielectric peaks occurring as a function of temperature, and shifting to higher temperatures as the frequency of the alternating current (a.c.) increases, are interpreted by Hemberger *et al.* as evidence that CdCr_2S_4 is a relaxor ferroelectric². In a relaxor, the temperature of the dielectric peak (T_m) depends on the a.c. frequency, f , in accordance with the Vogel–Fulcher law: $f = f_0 e^{-U/k_B(T_m - T_f)}$. Here, T_f is the finite temperature at which the relaxor dynamics ‘freeze’ into a polarizable state. We have fitted the data for CdCr_2S_4 (Fig. 1) and obtain an unrealistic value for T_f of about –175 K. For polarization to be remanent at the temperatures suggested by Hemberger *et al.*, T_f ought to be in the region of 60 K, which does not fit the data (Fig. 1).

Relaxor-like behaviour can actually appear in any heterogeneous semiconductor as a result of the Maxwell–Wagner effect, a conductive artefact unrelated to relaxors^{12,13}. The low-frequency divergence of the imaginary permittivity^{3–5} is indeed consistent with direct-current (d.c.) conductivity in a Maxwell–Wagner system¹³. Furthermore, any Maxwell–Wagner system (any heterogeneous semiconductor) that is magnetoresistive can show magnetocapacitance¹⁴ (Fig. 2), and the thiospinels are all magnetoresistive^{3–5}.

Other indicators of magnetocapacitance due to magnetoresistive artefacts¹⁴ are the large effect of magnetic fields on dielectric loss, the strong frequency dependence of the magnetocapacitance and its correlation with magnetoresistance^{2–5}. The magnetocapacitance in fact exceeds that possible for linear magnetoelectric coupling¹⁵ and, whereas larger magnetocapacitance is possible for nonlinear coupling, a different temperature and field dependence would be expected; in the thiospinels it is neither proportional to the magnetization M or M^2 , nor correlated with the alleged onset of polarization.

Hemberger *et al.* try to rule out contact artefacts by using different electrodes^{3–5}, but Maxwell–Wagner does not require contact-related depletion layers. The unusual dielectric features disappear on annealing⁴, indicating that there could be a heterogeneous distribution of defects that can be annealed out. The authors’ single crystals were grown with chlorine as a transport agent, so chlorine-based impurities are one possibility. The authors themselves acknowledge the possibil-

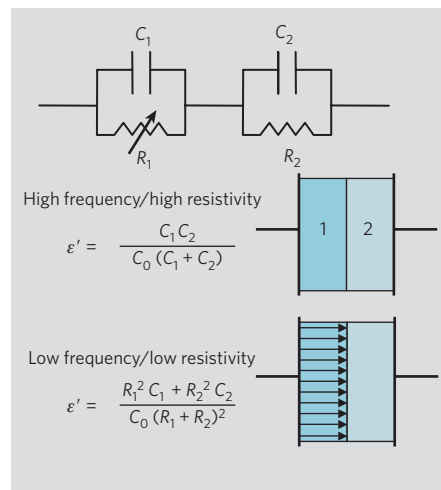


Figure 2 | Maxwell–Wagner equivalent circuit model of a heterogeneous lossy dielectric. The model shows approximated equations for the real part of the apparent dielectric constant (ϵ') at high and low frequencies of alternating current. At high frequencies, the two types of region act as two capacitors in series (C_1 and C_2). The effect of lowering the frequency is the same as that of decreasing the resistance (R) of one of the two components: it allows the current (arrows) to flow through, reducing the apparent thickness and increasing the measured capacitance (C). In the thiospinels, applying a magnetic field has the same effect on capacitance as reducing the frequency², consistent with negative magnetoresistance in a Maxwell–Wagner system¹⁴.

ity of sulphur non-stoichiometry¹⁰, so that is another possible culprit.

The greater message here is that magnetocapacitance is not a property of a material, but an impedance measurement of a circuit; it may reflect the intrinsic magnetodielectric properties of the crystal under study¹, or extrinsic parasitic contributions. Artificial magnetoelectric effects are common and do not require that the material be multiferroic. Because these artefacts are more likely in semiconducting samples, such as the spinels, and because all reported data are consistent with a magnetoresistive Maxwell–Wagner effect, this seems to be the most plausible explanation. At a minimum, our analysis shows that evidence for the spinels being true relaxor multiferroics is at present inconclusive.

Gustau Catalan, James F. Scott

Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK
e-mail: gcat05@esc.cam.ac.uk

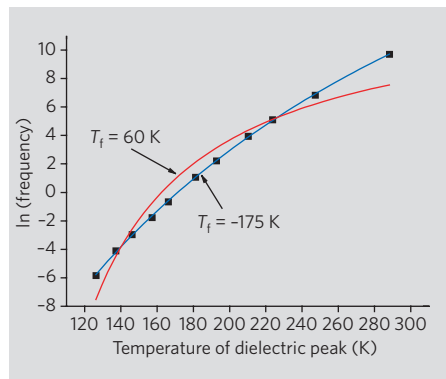


Figure 1 | Temperature of dielectric maxima as a function of alternating-current frequency. Data were extracted from refs 2–4 and fitted with the Vogel–Fulcher law of relaxors. The good fit (blue line) gives a negative value, whereas the T_f (the finite temperature at which the relaxor dynamics ‘freezes’ into a polarizable state) suggested by Hemberger *et al.*² (red line) does not reproduce the data. If T_f is not positive, there can be no remanent polarization. If an Arrhenius law is fitted to the dielectric peaks, the activation energy (about 0.33 eV) (refs 3–5) is similar to that obtained from fitting the conductivity data (0.2–0.3 eV), indicating that the dielectric relaxation is indeed due to the temperature dependence of the conductivity.

1. Eerenstein, W., Mathur, N. D. & Scott, J. F. Multiferroic and magnetoelectric materials. *Nature* **442**, 759–765 (2006).
2. Hemberger, J. *et al.* Relaxor ferroelectricity and colossal magnetocapacitive coupling in ferromagnetic CdCr_2S_4 .

- Nature* **434**, 364–367 (2005).
3. Lunkenheimer, P., Fichtl, R., Hemberger, J., Tsurkan, V. & Loidl, A. Relaxation dynamics and colossal magnetocapacitive effect in CdCr_2S_4 . *Phys. Rev. B* **72**, 60103 (2005).
 4. Hemberger, J. *et al.* Multiferroic behavior in CdCr_2X_4 (X = S, Se). arXiv:cond-mat/0508014 (2005).
 5. Weber, S. *et al.* Colossal magnetocapacitance and colossal magnetoresistance in HgCr_2S_4 . *Phys. Rev. Lett.* **96**, 157202 (2006).
 6. Lines, M. E. & Glass, A. M. *Principles and Applications of Ferroelectrics and Related Materials* (Clarendon, Oxford, 1977).
 7. Pintilie, L. & Alexe, M. Ferroelectric-like hysteresis loop in nonferroelectric systems. *Appl. Phys. Lett.* **87**, 112903 (2005).
 8. Fennie, C. J. & Rabe, K. M. Polar phonons and intrinsic dielectric response of the ferromagnetic insulating spinel CdCr_2S_4 from first principles. *Phys. Rev. B* **72**, 214123 (2005).
 9. Koshizuka, N., Ushiodo, S. & Tsushima, T. Resonance scattering in CdCr_2S_4 : magnetic-circular-polarization properties. *Phys. Rev. B* **21**, 1316–1322 (1980).
 10. Gnezdilov, V. *et al.* Evidence for local lattice distortions in giant magnetocapacitive CdCr_2S_4 . arXiv:cond-mat/0702362 (2007).
 11. Martin, R. M. & Damen, T. C. Breakdown of selection rules in resonance Raman scattering. *Phys. Rev. Lett.* **26**, 86–88 (1971).
 12. Catalan, G., O'Neill, D., Bowman, R. M. & Gregg, J. M. Relaxor features in ferroelectric superlattices: a Maxwell–Wagner approach. *Appl. Phys. Lett.* **77**, 3078–3080 (2000).
 13. O'Neill, D., Bowman, R. M. & Gregg, J. M. Dielectric enhancement and Maxwell–Wagner effects in ferroelectric superlattice structures. *Appl. Phys. Lett.* **77**, 1520–1522 (2000).
 14. Catalan, G. Magnetocapacitance without magnetoelectric coupling. *Appl. Phys. Lett.* **88**, 102902 (2006).
 15. Brown, W. F. Jr, Hornreich, R. M. & Shtrikman, S. Upper bound on the magnetoelectric susceptibility. *Phys. Rev.* **168**, 574–577 (1968).

Received 29 January; accepted 10 July 2007.

Competing financial interests: declared none.

doi:10.1038/nature06156

MAGNETOELECTRICS

Hemberger *et al.* reply

Replying to: G. Catalan & J. F. Scott *Nature* **448**, doi:10.1038/nature06156 (2007)

Catalan and Scott¹ propose an alternative interpretation for our findings² for CdCr_2S_4 in terms of Maxwell–Wagner effects. They also quote related isomorphs, such as HgCr_2S_4 (ref. 3), which has been discussed elsewhere^{4,5}. As we have shown, Maxwell–Wagner relaxations can indeed strongly affect the dielectric properties of transition-metal oxides^{6,7} and semiconductors⁸. We do not find Catalan and Scott's arguments about CdCr_2S_4 convincing, however, and think that our experiments provide ample evidence for multiferroic relaxor behaviour.

Taking the arguments of Catalan and Scott point by point, we explain why we stand by our interpretation. First, lossy dielectrics reveal purely elliptical hysteresis loops. The observed hysteresis is typical for relaxors when saturation is not reached⁹. CdCr_2S_4 is not a proper ferroelectric. Because of the history and time dependence of relaxors⁹, the thermo-remnant polarization cannot be compared with results from $P(E)$ loops. We excluded space-charge effects by frequency-dependent $P(E)$ measurements³.

Second, coupled spin and polarization waves, not soft phonons, are the relevant collective excitations in multiferroics^{10,11}. Recent Raman experiments¹² revealing local polar distortions cannot be explained by resonance effects: those data were determined with energy that is much larger than the red-shifted temperature-dependent interband excitation. Consequently, the effect of local distortions on the longitudinal-optical intensity is larger than possible resonance contributions. Furthermore, the different temperature dependencies of the longitudinal-optical mode intensities, the absent multiphonon and fluorescent contributions, and small phonon linewidth speak against strong resonance contributions.

In their Fig. 1, Catalan and Scott analyse the peak positions of the temperature-dependent

dielectric constant, ϵ' , reported by us^{2,13}. Such data do not provide precise information on relaxation dynamics and are hampered by conductivity contributions in semiconducting CdCr_2S_4 . Figure 1 presents more accurate information on the relaxation rate¹³. At high temperatures, a reasonable description with $T_f \approx 60$ K is possible. An analysis based on $\epsilon''(T)$, as performed by Catalan and Scott¹, and including the transition regime at low temperatures, leads to an unrealistic value for T_f .

The low-frequency divergence of the loss is a common feature of materials with significant conductivity. The circuit discussed by Catalan and Scott¹ cannot account for the complex behaviour of CdCr_2S_4 and more elaborate networks^{6,7} would be needed. Magnetoresistance is a necessary but insufficient condition for Maxwell–Wagner-related magnetocapacitance. The field changes the relaxation dynamics^{2,13}, which explains the effect on the relaxation-related loss and the frequency-dependent magnetocapacitance. The correlation between magnetocapacitance and magnetoresistance arises because both are triggered by the strong field-induced shift of the ferromagnetic transition temperature. Concerning the magnitude of the magnetocapacitance, CdCr_2S_4 is not a conventional multiferroic and cannot be treated with the same theoretical footing². The dynamic nature of the effect^{2,13} explains the lack of correlation between polarization and magnetocapacitance.

From electron probe microanalysis and the narrow linewidths in X-ray studies using synchrotron radiation, we exclude any non-homogeneous impurity distribution in our single crystals. Meanwhile, we have reproduced all effects in annealed single crystals grown without chlorine and in ceramics doped with indium.

We therefore think that our experiments provide evidence for intrinsic multiferroic

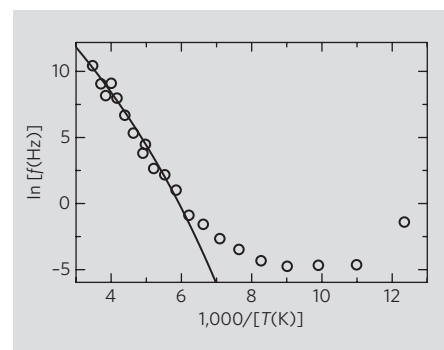


Figure 1 | Temperature dependence of the relaxation rate of CdCr_2S_4 . The relaxation rate, f , was calculated from the relaxation times τ , shown in ref. 13, determined from fits of the frequency-dependent complex permittivity. These data are more accurate than those analysed in Fig. 1 of ref. 1. As revealed by the Arrhenius representation of Fig. 1, at high temperatures the data follow thermally activated behaviour¹³, but they also can be fitted by a Vogel–Fulcher law, with $T_f = 60$ K (line). At low temperatures, strong ferromagnetic fluctuations² lead to deviations owing to an acceleration of the relaxation dynamics^{2,13}.

behaviour, supporting early assumptions of polar phase transitions and magnetoelectric effects in spinels¹⁴.

Joachim Hemberger*, **Peter Lunkenheimer***, **Robert Fichtl***, **Hans-Albrecht Krug von Nidda***, **Vladimir Tsurkan***†, **Alois Loidl***

*Experimental Physics V, Centre for Electronic Correlations and Magnetism, University of Augsburg, 86159 Augsburg, Germany
e-mail: peter.lunkenheimer@physik.uni-augsburg.de

†Institute for Applied Physics, Academy of Sciences of Moldova, MD-2028, Chisinau, Moldova

1. Catalan, G. & Scott, J. F. *Nature* **448**, doi:10.1038/nature06156 (2007).

2. Hemberger, J. *et al.* Relaxor ferroelectricity and colossal

- magnetocapacitive coupling in ferromagnetic CdCr_2S_4 . *Nature* **434**, 364–367 (2005).
3. Weber, S. *et al.* Colossal magnetocapacitance and colossal magnetoresistance in HgCr_2S_4 . *Phys. Rev. Lett.* **96**, 157202 (2006).
 4. Catalan, G. & Scott, J. F. Comments on “Giant dielectric response in the one-dimensional charge-ordered semiconductor $(\text{NbSe}_4)_3\text{I}$ ” and “Colossal magnetocapacitance and colossal magnetoresistance in HgCr_2S_4 ”. Preprint at arXiv:cond-mat/0607500 (2006).
 5. Lunkenheimer, P. *et al.* Reply to comments on “Giant dielectric response in the one-dimensional charge-ordered semiconductor $(\text{NbSe}_4)_3\text{I}$ ” and “Colossal magnetocapacitance and colossal magnetoresistance in HgCr_2S_4 ”. Preprint at arXiv:cond-mat/0701417 (2007).
 6. Lunkenheimer, P., Fichtl, R., Ebbinghaus, S. G. & Loidl, A. Nonintrinsic origin of the colossal dielectric constants in $\text{CaCu}_3\text{Ti}_4\text{O}_{12}$. *Phys. Rev. B* **70**, 172102 (2004).
 7. Bobnar, V., Lunkenheimer, P., Paraskevopoulos, M. & Loidl, A. Separation of grain boundary effects and intrinsic properties in perovskite-like $\text{Gd}_{0.6}\text{Y}_{0.4}\text{BaCo}_2\text{O}_{5.5}$ using high-frequency dielectric spectroscopy. *Phys. Rev. B* **65**, 184403 (2002).
 8. Ritus, A. I. *et al.* Determination of the parameters of semiconducting $\text{CdF}_2\text{:In}$ with Schottky barriers from radio-frequency measurements. *Phys. Rev. B* **65**, 165209 (2002).
 9. Samara, G. A. The relaxational properties of compositionally disordered ABO_3 perovskites. *J. Phys.: Condens. Matter* **15**, R367–R411 (2003).
 10. Pimenov, A. *et al.* Possible evidence for electromagnons in multiferroic manganites. *Nature Phys.* **2**, 97–100 (2006).
 11. Katsura, H., Balatsky, A. V. & Nagaosa, N. Dynamical magnetoelectric coupling in helical magnets. *Phys. Rev. Lett.* **98**, 027203 (2007).
 12. Gnezdilov, V. *et al.* Evidence for local lattice distortions in giant magnetocapacitive CdCr_2S_4 . Preprint at arXiv:cond-mat/0702362 (2007).
 13. Lunkenheimer, P., Fichtl, R., Hemberger, J., Tsurkan, V. & Loidl, A. Relaxation dynamics and colossal magnetocapacitive effect in CdCr_2S_4 . *Phys. Rev. B* **72**, 060103 (2005).
 14. Schmid, H. & Ascher, E. Are antiferroelectricity and other physical properties ‘hidden’ in spinel compounds? *J. Phys. C: Solid State Phys.* **7**, 2697–2706 (1974).

doi:10.1038/nature06157

MAGNETOELECTRICS

Is CdCr_2S_4 a multiferroic relaxor?Arising from: J. Hemberger *et al.* *Nature* **434**, 364–367 (2005)

Materials showing simultaneous ferroelectric and magnetic ordering are attracting a great deal of interest because of their unusual physics and potential applications¹. Hemberger *et al.*² have reported relaxor-like dielectric properties and colossal magnetocapacitance (in excess of 500%) for the cubic spinel compound CdCr_2S_4 and related isomorphs^{3–5}, concluding that CdCr_2S_4 is a multiferroic relaxor. We argue here, however, that their results might also be explained by a conductive artefact.

The polarization hysteresis loops of Hemberger *et al.*² have a shape unlike that seen in ferroelectrics and more like that of a lossy dielectric^{6,7}, such as the semiconducting thiospinels. The ‘remanent’ polarization is much larger than the pyroelectric one, whereas switchable polarization must always be smaller than the pyroelectric polarization — which on its own calls the hysteresis results into question. The pyroelectric polarizations of 0.05–0.07 $\mu\text{C cm}^{-2}$ are themselves tiny for any proper ferroelectric, and are consistent with thermally stimulated discharge of space-charge injected during poling⁶.

Ab initio calculations⁸ preclude soft phonons, and hence proper ferroelectricity. Nor have Raman experiments⁹ found soft phonons or a lack of centrosymmetry. Raman results with wavelengths near or below the resonant edge¹⁰ are consistent with local non-centro-

symmetry (a necessary, although insufficient, condition for ferroelectricity) — however, this technique is unreliable because nominally forbidden odd-parity longitudinal-optical phonon scattering can be strong even in centric crystals¹¹ when there is real absorption.

Broad dielectric peaks occurring as a function of temperature, and shifting to higher temperatures as the frequency of the alternating current (a.c.) increases, are interpreted by Hemberger *et al.* as evidence that CdCr_2S_4 is a relaxor ferroelectric². In a relaxor, the temperature of the dielectric peak (T_m) depends on the a.c. frequency, f , in accordance with the Vogel–Fulcher law: $f = f_0 e^{-U/k_B(T_m - T_f)}$. Here, T_f is the finite temperature at which the relaxor dynamics ‘freeze’ into a polarizable state. We have fitted the data for CdCr_2S_4 (Fig. 1) and obtain an unrealistic value for T_f of about –175 K. For polarization to be remanent at the temperatures suggested by Hemberger *et al.*, T_f ought to be in the region of 60 K, which does not fit the data (Fig. 1).

Relaxor-like behaviour can actually appear in any heterogeneous semiconductor as a result of the Maxwell–Wagner effect, a conductive artefact unrelated to relaxors^{12,13}. The low-frequency divergence of the imaginary permittivity^{3–5} is indeed consistent with direct-current (d.c.) conductivity in a Maxwell–Wagner system¹³. Furthermore, any Maxwell–Wagner system (any heterogeneous semiconductor) that is magnetoresistive can show magnetocapacitance¹⁴ (Fig. 2), and the thiospinels are all magnetoresistive^{3–5}.

Other indicators of magnetocapacitance due to magnetoresistive artefacts¹⁴ are the large effect of magnetic fields on dielectric loss, the strong frequency dependence of the magnetocapacitance and its correlation with magnetoresistance^{2–5}. The magnetocapacitance in fact exceeds that possible for linear magnetoelectric coupling¹⁵ and, whereas larger magnetocapacitance is possible for nonlinear coupling, a different temperature and field dependence would be expected; in the thiospinels it is neither proportional to the magnetization M or M^2 , nor correlated with the alleged onset of polarization.

Hemberger *et al.* try to rule out contact artefacts by using different electrodes^{3–5}, but Maxwell–Wagner does not require contact-related depletion layers. The unusual dielectric features disappear on annealing⁴, indicating that there could be a heterogeneous distribution of defects that can be annealed out. The authors’ single crystals were grown with chlorine as a transport agent, so chlorine-based impurities are one possibility. The authors themselves acknowledge the possibil-

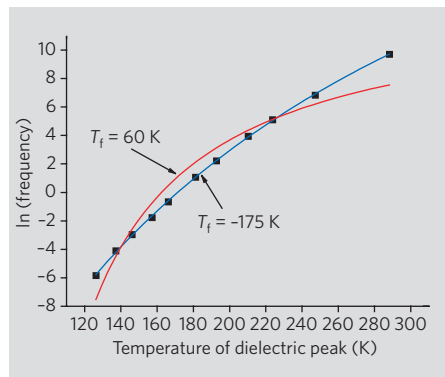


Figure 1 | Temperature of dielectric maxima as a function of alternating-current frequency. Data were extracted from refs 2–4 and fitted with the Vogel–Fulcher law of relaxors. The good fit (blue line) gives a negative value, whereas the T_f (the finite temperature at which the relaxor dynamics ‘freezes’ into a polarizable state) suggested by Hemberger *et al.*² (red line) does not reproduce the data. If T_f is not positive, there can be no remanent polarization. If an Arrhenius law is fitted to the dielectric peaks, the activation energy (about 0.33 eV) (refs 3–5) is similar to that obtained from fitting the conductivity data (0.2–0.3 eV), indicating that the dielectric relaxation is indeed due to the temperature dependence of the conductivity.

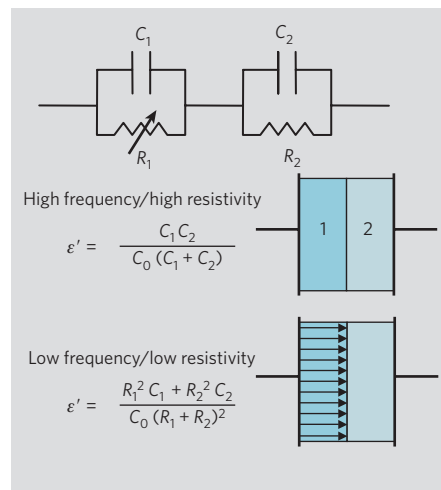


Figure 2 | Maxwell–Wagner equivalent circuit model of a heterogeneous lossy dielectric. The model shows approximated equations for the real part of the apparent dielectric constant (ϵ') at high and low frequencies of alternating current. At high frequencies, the two types of region act as two capacitors in series (C_1 and C_2). The effect of lowering the frequency is the same as that of decreasing the resistance (R) of one of the two components: it allows the current (arrows) to flow through, reducing the apparent thickness and increasing the measured capacitance (C). In the thiospinels, applying a magnetic field has the same effect on capacitance as reducing the frequency², consistent with negative magnetoresistance in a Maxwell–Wagner system¹⁴.

ity of sulphur non-stoichiometry¹⁰, so that is another possible culprit.

The greater message here is that magnetocapacitance is not a property of a material, but an impedance measurement of a circuit; it may reflect the intrinsic magnetodielectric properties of the crystal under study¹, or extrinsic parasitic contributions. Artificial magnetoelectric effects are common and do not require that the material be multiferroic. Because these artefacts are more likely in semiconducting samples, such as the spinels, and because all reported data are consistent with a magnetoresistive Maxwell–Wagner effect, this seems to be the most plausible explanation. At a minimum, our analysis shows that evidence for the spinels being true relaxor multiferroics is at present inconclusive.

Gustau Catalan, James F. Scott

Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK
e-mail: gcat05@esc.cam.ac.uk

1. Eerenstein, W., Mathur, N. D. & Scott, J. F. Multiferroic and magnetoelectric materials. *Nature* **442**, 759–765 (2006).
2. Hemberger, J. *et al.* Relaxor ferroelectricity and colossal magnetocapacitive coupling in ferromagnetic CdCr_2S_4 .

- Nature* **434**, 364–367 (2005).
3. Lunkenheimer, P., Fichtl, R., Hemberger, J., Tsurkan, V. & Loidl, A. Relaxation dynamics and colossal magnetocapacitive effect in CdCr_2S_4 . *Phys. Rev. B* **72**, 60103 (2005).
 4. Hemberger, J. *et al.* Multiferroic behavior in CdCr_2X_4 (X = S, Se). arXiv:cond-mat/0508014 (2005).
 5. Weber, S. *et al.* Colossal magnetocapacitance and colossal magnetoresistance in HgCr_2S_4 . *Phys. Rev. Lett.* **96**, 157202 (2006).
 6. Lines, M. E. & Glass, A. M. *Principles and Applications of Ferroelectrics and Related Materials* (Clarendon, Oxford, 1977).
 7. Pintilie, L. & Alexe, M. Ferroelectric-like hysteresis loop in nonferroelectric systems. *Appl. Phys. Lett.* **87**, 112903 (2005).
 8. Fennie, C. J. & Rabe, K. M. Polar phonons and intrinsic dielectric response of the ferromagnetic insulating spinel CdCr_2S_4 from first principles. *Phys. Rev. B* **72**, 214123 (2005).
 9. Koshizuka, N., Ushiodo, S. & Tsushima, T. Resonance scattering in CdCr_2S_4 : magnetic-circular-polarization properties. *Phys. Rev. B* **21**, 1316–1322 (1980).
 10. Gnezdilov, V. *et al.* Evidence for local lattice distortions in giant magnetocapacitive CdCr_2S_4 . arXiv:cond-mat/0702362 (2007).
 11. Martin, R. M. & Damen, T. C. Breakdown of selection rules in resonance Raman scattering. *Phys. Rev. Lett.* **26**, 86–88 (1971).
 12. Catalan, G., O'Neill, D., Bowman, R. M. & Gregg, J. M. Relaxor features in ferroelectric superlattices: a Maxwell–Wagner approach. *Appl. Phys. Lett.* **77**, 3078–3080 (2000).
 13. O'Neill, D., Bowman, R. M. & Gregg, J. M. Dielectric enhancement and Maxwell–Wagner effects in ferroelectric superlattice structures. *Appl. Phys. Lett.* **77**, 1520–1522 (2000).
 14. Catalan, G. Magnetocapacitance without magnetoelectric coupling. *Appl. Phys. Lett.* **88**, 102902 (2006).
 15. Brown, W. F. Jr, Hornreich, R. M. & Shtrikman, S. Upper bound on the magnetoelectric susceptibility. *Phys. Rev.* **168**, 574–577 (1968).

Received 29 January; accepted 10 July 2007.

Competing financial interests: declared none.

doi:10.1038/nature06156

MAGNETOELECTRICS

Hemberger *et al.* reply

Replying to: G. Catalan & J. F. Scott *Nature* **448**, doi:10.1038/nature06156 (2007)

Catalan and Scott¹ propose an alternative interpretation for our findings² for CdCr_2S_4 in terms of Maxwell–Wagner effects. They also quote related isomorphs, such as HgCr_2S_4 (ref. 3), which has been discussed elsewhere^{4,5}. As we have shown, Maxwell–Wagner relaxations can indeed strongly affect the dielectric properties of transition-metal oxides^{6,7} and semiconductors⁸. We do not find Catalan and Scott's arguments about CdCr_2S_4 convincing, however, and think that our experiments provide ample evidence for multiferroic relaxor behaviour.

Taking the arguments of Catalan and Scott point by point, we explain why we stand by our interpretation. First, lossy dielectrics reveal purely elliptical hysteresis loops. The observed hysteresis is typical for relaxors when saturation is not reached⁹. CdCr_2S_4 is not a proper ferroelectric. Because of the history and time dependence of relaxors⁹, the thermo-remnant polarization cannot be compared with results from $P(E)$ loops. We excluded space-charge effects by frequency-dependent $P(E)$ measurements³.

Second, coupled spin and polarization waves, not soft phonons, are the relevant collective excitations in multiferroics^{10,11}. Recent Raman experiments¹² revealing local polar distortions cannot be explained by resonance effects: those data were determined with energy that is much larger than the red-shifted temperature-dependent interband excitation. Consequently, the effect of local distortions on the longitudinal-optical intensity is larger than possible resonance contributions. Furthermore, the different temperature dependencies of the longitudinal-optical mode intensities, the absent multiphonon and fluorescent contributions, and small phonon linewidth speak against strong resonance contributions.

In their Fig. 1, Catalan and Scott analyse the peak positions of the temperature-dependent

dielectric constant, ϵ' , reported by us^{2,13}. Such data do not provide precise information on relaxation dynamics and are hampered by conductivity contributions in semiconducting CdCr_2S_4 . Figure 1 presents more accurate information on the relaxation rate¹³. At high temperatures, a reasonable description with $T_f \approx 60$ K is possible. An analysis based on $\epsilon''(T)$, as performed by Catalan and Scott¹, and including the transition regime at low temperatures, leads to an unrealistic value for T_f .

The low-frequency divergence of the loss is a common feature of materials with significant conductivity. The circuit discussed by Catalan and Scott¹ cannot account for the complex behaviour of CdCr_2S_4 and more elaborate networks^{6,7} would be needed. Magnetoresistance is a necessary but insufficient condition for Maxwell–Wagner-related magnetocapacitance. The field changes the relaxation dynamics^{2,13}, which explains the effect on the relaxation-related loss and the frequency-dependent magnetocapacitance. The correlation between magnetocapacitance and magnetoresistance arises because both are triggered by the strong field-induced shift of the ferromagnetic transition temperature. Concerning the magnitude of the magnetocapacitance, CdCr_2S_4 is not a conventional multiferroic and cannot be treated with the same theoretical footing². The dynamic nature of the effect^{2,13} explains the lack of correlation between polarization and magnetocapacitance.

From electron probe microanalysis and the narrow linewidths in X-ray studies using synchrotron radiation, we exclude any non-homogeneous impurity distribution in our single crystals. Meanwhile, we have reproduced all effects in annealed single crystals grown without chlorine and in ceramics doped with indium.

We therefore think that our experiments provide evidence for intrinsic multiferroic

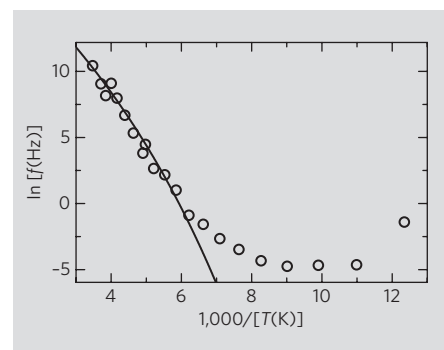


Figure 1 | Temperature dependence of the relaxation rate of CdCr_2S_4 . The relaxation rate, f , was calculated from the relaxation times τ , shown in ref. 13, determined from fits of the frequency-dependent complex permittivity. These data are more accurate than those analysed in Fig. 1 of ref. 1. As revealed by the Arrhenius representation of Fig. 1, at high temperatures the data follow thermally activated behaviour¹³, but they also can be fitted by a Vogel–Fulcher law, with $T_f = 60$ K (line). At low temperatures, strong ferromagnetic fluctuations² lead to deviations owing to an acceleration of the relaxation dynamics^{2,13}.

behaviour, supporting early assumptions of polar phase transitions and magnetoelectric effects in spinels¹⁴.

Joachim Hemberger*, **Peter Lunkenheimer***, **Robert Fichtl***, **Hans-Albrecht Krug von Nidda***, **Vladimir Tsurkan*†**, **Alois Loidl***

*Experimental Physics V, Centre for Electronic Correlations and Magnetism, University of Augsburg, 86159 Augsburg, Germany
e-mail: peter.lunkenheimer@physik.uni-augsburg.de

†Institute for Applied Physics, Academy of Sciences of Moldova, MD-2028, Chisinau, Moldova

1. Catalan, G. & Scott, J. F. *Nature* **448**, doi:10.1038/nature06156 (2007).

2. Hemberger, J. *et al.* Relaxor ferroelectricity and colossal

- magnetocapacitive coupling in ferromagnetic CdCr_2S_4 . *Nature* **434**, 364–367 (2005).
3. Weber, S. *et al.* Colossal magnetocapacitance and colossal magnetoresistance in HgCr_2S_4 . *Phys. Rev. Lett.* **96**, 157202 (2006).
 4. Catalan, G. & Scott, J. F. Comments on “Giant dielectric response in the one-dimensional charge-ordered semiconductor $(\text{NbSe}_4)_3\text{I}$ ” and “Colossal magnetocapacitance and colossal magnetoresistance in HgCr_2S_4 ”. Preprint at arXiv:cond-mat/0607500 (2006).
 5. Lunkenheimer, P. *et al.* Reply to comments on “Giant dielectric response in the one-dimensional charge-ordered semiconductor $(\text{NbSe}_4)_3\text{I}$ ” and “Colossal magnetocapacitance and colossal magnetoresistance in HgCr_2S_4 ”. Preprint at arXiv:cond-mat/0701417 (2007).
 6. Lunkenheimer, P., Fichtl, R., Ebbinghaus, S. G. & Loidl, A. Nonintrinsic origin of the colossal dielectric constants in $\text{CaCu}_3\text{Ti}_4\text{O}_{12}$. *Phys. Rev. B* **70**, 172102 (2004).
 7. Bobnar, V., Lunkenheimer, P., Paraskevopoulos, M. & Loidl, A. Separation of grain boundary effects and intrinsic properties in perovskite-like $\text{Gd}_{0.6}\text{Y}_{0.4}\text{BaCo}_2\text{O}_{5.5}$ using high-frequency dielectric spectroscopy. *Phys. Rev. B* **65**, 184403 (2002).
 8. Ritus, A. I. *et al.* Determination of the parameters of semiconducting $\text{CdF}_2\text{:In}$ with Schottky barriers from radio-frequency measurements. *Phys. Rev. B* **65**, 165209 (2002).
 9. Samara, G. A. The relaxational properties of compositionally disordered ABO_3 perovskites. *J. Phys.: Condens. Matter* **15**, R367–R411 (2003).
 10. Pimenov, A. *et al.* Possible evidence for electromagnons in multiferroic manganites. *Nature Phys.* **2**, 97–100 (2006).
 11. Katsura, H., Balatsky, A. V. & Nagaosa, N. Dynamical magnetoelectric coupling in helical magnets. *Phys. Rev. Lett.* **98**, 027203 (2007).
 12. Gnezdilov, V. *et al.* Evidence for local lattice distortions in giant magnetocapacitive CdCr_2S_4 . Preprint at arXiv:cond-mat/0702362 (2007).
 13. Lunkenheimer, P., Fichtl, R., Hemberger, J., Tsurkan, V. & Loidl, A. Relaxation dynamics and colossal magnetocapacitive effect in CdCr_2S_4 . *Phys. Rev. B* **72**, 060103 (2005).
 14. Schmid, H. & Ascher, E. Are antiferroelectricity and other physical properties ‘hidden’ in spinel compounds? *J. Phys. C: Solid State Phys.* **7**, 2697–2706 (1974).

doi:10.1038/nature06157

Progressive field-state collapse and quantum non-demolition photon counting

Christine Guerlin¹, Julien Bernu¹, Samuel Deléglise¹, Clément Sayrin¹, Sébastien Gleyzes¹, Stefan Kuhr¹†, Michel Brune¹, Jean-Michel Raimond¹ & Serge Haroche^{1,2}

The irreversible evolution of a microscopic system under measurement is a central feature of quantum theory. From an initial state generally exhibiting quantum uncertainty in the measured observable, the system is projected into a state in which this observable becomes precisely known. Its value is random, with a probability determined by the initial system's state. The evolution induced by measurement (known as 'state collapse') can be progressive, accumulating the effects of elementary state changes. Here we report the observation of such a step-by-step collapse by non-destructively measuring the photon number of a field stored in a cavity. Atoms behaving as microscopic clocks cross the cavity successively. By measuring the light-induced alterations of the clock rate, information is progressively extracted, until the initially uncertain photon number converges to an integer. The suppression of the photon number spread is demonstrated by correlations between repeated measurements. The procedure illustrates all the postulates of quantum measurement (state collapse, statistical results and repeatability) and should facilitate studies of non-classical fields trapped in cavities.

The projection of a microscopic system into an eigenstate of the measured observable reflects the change of knowledge produced by the measurement. Information is either acquired in a single step, as in a Stern–Gerlach spin-component measurement¹, or in an incremental way, as in spin-squeezing experiments^{2,3}. A projective measurement is called 'quantum non-demolition'^{4–8} (QND) when the collapsed state is invariant under the system's free unitary evolution. Sequences of repeated measurements then yield identical results and jumps between different outcomes reveal an external perturbation^{7,8}.

Various QND measurements have been realized on massive particles. The motional state of a trapped electron has been measured through the current induced in the trapping electrodes⁹. The internal state of trapped ions has been read out, directly by way of laser-induced fluorescence¹⁰, or indirectly through quantum gate operations entangling them to an ancillary ion¹¹. Collective spin states of an atomic ensemble have been QND-detected through its dispersive interaction with light¹².

QND light measurements are especially challenging, as photons are detected with photosensitive materials that usually absorb them. Photon demolition is however avoidable¹³. In non-resonant processes, light induces nonlinear dispersive effects¹⁴ in a medium, without real transitions. Photons can then be detected without loss. Dispersive schemes have been applied to detect the fluctuations of a signal light beam by the phase shifts it induces on a probe beam interacting with the same medium^{15,16}. Neither these methods, nor alternative ones based on the noiseless duplication of light by optical parametric amplifiers^{17,18}, have been able, so far, to pin down photon numbers.

Single-photon resolution requires an extremely strong light–matter coupling, optimally achieved by confining radiation inside a cavity. This is the domain of cavity quantum electrodynamics^{19–21}, in which experiments attaining single-quantum resolution have been performed with optical^{22,23} or microwave photons, the latter being coupled either to Rydberg atoms^{24–26} or to superconducting junctions²⁷. In a QND experiment, cavity losses should be negligible

during a sequence of repeated measurements. We have realized a superconducting cavity with a very long field damping time²⁸, and used it to detect repeatedly a single photon²⁹. Here, we demonstrate with this cavity a general QND photon counting method applied to a microwave field containing several photons. It implements a variant of a procedure proposed in refs 30 and 31, and illustrates all the postulates of a projective measurement¹.

A stream of atoms crosses the cavity and performs a step-by-step measurement of the photon number-dependent alteration of the atomic transition frequency known as the 'light shift'. We follow the measurement-induced evolution from a coherent state of light into a Fock state of well-defined energy, containing up to 7 photons. Repeating the measurement on the collapsed state yields the same result, until cavity damping makes the photon number decrease. The measured field energy then decays by quantum jumps along a staircase-like cascade, ending in vacuum.

In this experiment, light is an object of investigation repeatedly interrogated by atoms. Its evolution under continuous non-destructive monitoring is directly accessible to measurement, making real the stochastic trajectories of quantum field Monte Carlo simulations^{20,32}. Repeatedly counting photons in a cavity as marbles in a box opens novel perspectives for studying non-classical states of radiation.

An atomic clock to count photons

To explain our QND method, consider the thought experiment sketched in Fig. 1a. A photon box, similar to the contraction imagined in another context by Einstein and Bohr¹, contains a few photons together with a clock whose rate is affected by the light. Depending upon the photon number n , the hand of the clock points in different directions after a given interaction time with the field. This time is set so that a photon causes a π/q angular shift of the hand (here q is an integer). There are $2q$ values (0, 1, ..., $2q-1$) of the photon number corresponding to regularly spaced directions of the hand, spanning 360° (Fig. 1a shows the hand's positions for $q = 4$ and

¹Laboratoire Kastler Brossel, Ecole Normale Supérieure, CNRS, Université Pierre et Marie Curie, 24 rue Lhomond, 75231 Paris Cedex 05, France. ²Collège de France, 11 place Marcelin Berthelot, 75231 Paris Cedex 05, France. †Present address: Johannes Gutenberg Universität, Institut für Physik, Staudingerweg 7, 55128 Mainz, Germany.

$n = 0, 1, 3$). For $n \geq 2q$, the readings periodically repeat and the clock measures n modulo $2q$.

This description is translatable into the atomic world (Fig. 1b). The evolution of a two-level atom crossing the cavity C is described as the rotation of a spin evolving on a Bloch sphere³³. The atomic levels, $|\pm_z\rangle$, correspond to up and down spin states along the O–z direction. Before entering C, the spin is rotated by a pulse R_1 from $|+_z\rangle$ into state $|+_x\rangle = (|+_z\rangle + |-_z\rangle)/\sqrt{2}$, represented by a vector along O–x. This vector then starts to rotate in the x–O–y equatorial plane, in analogy with the ticking of the hand in the thought experiment. The atomic flight time across C is adjusted to result, per photon, in a π/q light-induced rotation of the spin. The $2q$ final spin states $|+_n\rangle$ correlated with $0 \leq n < 2q$ correspond to vectors reproducing the positions of the classical clock's hand.

In general, the photon number exhibits quantum uncertainty. The field, initially in a superposition $\sum_n c_n |n\rangle$ of Fock states $|n\rangle$, gets entangled with the spin, the final atom–field state becoming $\sum_n c_n |+_n\rangle \otimes |n\rangle$. The spin points in a fuzzy direction that the QND measurement is designed to pin down. As an example, consider a field in a coherent state³⁴ of complex amplitude α defined by the C-numbers $c_n = \exp(-|\alpha|^2/2)(\alpha^n/\sqrt{n!})$. Its photon number distribution, $P_0(n) = |c_n|^2$, is poissonian with an average photon number $n_0 = |\alpha|^2$ and a spread $\Delta n = \sqrt{n_0}$.

If we could determine the final atom state, the clock's delay—and hence n —would be read in a single measurement. This is however forbidden by quantum theory³⁵. The $2q$ spin states are not mutually orthogonal (except for $q = 1$, see below) and cannot be unambiguously distinguished (this ambiguity is exploited in quantum cryptography³⁶). Only partial information can be extracted from a spin, namely its projection along a direction O–u in the x–O–y plane making an arbitrary angle ϕ with O–x.

The angle between O–u and the direction of the $|+_n\rangle$ spin state is $n\pi/q - \phi$. The conditional probabilities for detecting the $|\pm_u\rangle$ states when C contains n photons are $P(j, \phi|n) = [1 + \cos(n\pi/q - \phi + j\pi)]/2$ (using quantum information notation³³, we assign to the \pm spin states the values $j = 0/1$ and rename $|j, \phi\rangle$ the states $|\pm_u\rangle$). Measuring the spin along O–u is performed by submitting it, after cavity exit, to a pulse R_2 whose phase is set to map O–u onto O–z (Fig. 1b). This rotation is followed by the measurement of the atom's energy, equivalent to a spin detection along O–z. The combination of R_1 and R_2 is a Ramsey interferometer³⁷. The probabilities for finding $j = 0$ and 1

along O–z oscillate versus ϕ , which is a typical feature of quantum interference (Methods).

The $q = 1$ case is a notable exception for which a single measurement yields complete information. There are then only two opposite hand positions on the atomic clock dial, corresponding to orthogonal states. Ideally, a single detection pins down n modulo 2, yielding the photon number parity. For weak fields with $n_0 \ll 1$, the probability for $n > 1$ is negligible and the parity defines n . The telegraphic signals²⁹ obtained by detecting a stream of atoms reveal the photon number evolution, with quantum jumps between $n = 0$ and 1 as the field randomly exchanges energy with the cavity walls. For larger fields, though, information must be extracted in a subtler way.

Progressive pinning-down of photon number

The random outcome of a spin detection modifies our knowledge of the photon number distribution. The conditional probability $P(n|j, \phi)$ for finding n photons after detecting the spin value j along O–u is related to the inverse conditional probability $P(j, \phi|n)$ by Bayes' law³⁸:

$$P(n|j, \phi) = P_0(n)P(j, \phi|n)/P(j, \phi) = P_0(n)[1 + \cos(n\pi/q - \phi + j\pi)]/2P(j, \phi) \quad (1)$$

where $P(j, \phi) = \sum_n P(j, \phi|n)P_0(n)$ is the a priori probability for j . This formula directly follows from the definition of conditional probabilities. It can also be derived from the projection postulate¹. After detection of the spin in state $|j, \phi\rangle$, the entangled atom–field system collapses into $[\sum_n c_n \langle j, \phi|+_n\rangle |n\rangle] \otimes |j, \phi\rangle/\sqrt{P(j, \phi)}$. This entails that the photon number probability is (up to a global factor) multiplied by $|\langle j, \phi|+_n\rangle|^2 = P(j, \phi|n)$.

Equation (1) embodies the logic of our QND procedure. The spin measurement has the effect of multiplying $P_0(n)$ by $P(j, \phi|n)$, which is a periodic function vanishing for specific values of n when ϕ is properly adjusted. If we choose $\phi = \pi p/q$ (where p is an integer), O–u points along the direction of $|+_p\rangle$. This entails $P(j = 1, \phi|p) = P(j = 0, \phi|p + q) = 0$. Detecting the spin in 1 (resp. 0) excludes the photon number $n = p$ (resp. $p + q$) as these outcomes are forbidden for the corresponding Fock states. One of the probabilities for finding $n = p$ or $n = p + q$ is cancelled, while the other is enhanced (when normalization is accounted for). At the same time, the probabilities of other photon numbers are modified according to equation (1).

This decimation is robust against imperfections. In a realistic situation, the theoretical probability $P(j, \phi|n)$ becomes $P^{(\text{exp})}(j, \phi|n) = [A + B \cos(n\Phi - \phi + j\pi)]/2$, where A and B are the Ramsey interferometer fringes offset and contrast, somewhat different from 1. The phase shift per photon Φ may also slightly depart from π/q . Before a QND measurement, A , B and Φ are determined by independent calibration. The limited contrast of the interferometer corresponds to a statistical uncertainty in the final atomic state. The atom and the field must then be described by density operators instead of pure states. The formula (1) remains valid with $P(j, \phi|n)$ replaced by $P^{(\text{exp})}(j, \phi|n)$. This is justified by Bayes' law or by generalization of the measurement postulate to statistical mixtures³³.

In order to obtain more information, we repeat the process and send a sequence of atoms across C. This results in a step-by-step change of the photon number distribution. From one atom to the next, we vary ϕ . Calling $\phi(k)$ the detection angle for the k th atom and $j(k)$ its spin reading, the photon number distribution after N atoms becomes:

$$P_N(n) = \frac{P_0(n)}{Z} \prod_{k=1}^N [A + B \cos(n\Phi - \phi(k) + j(k)\pi)] \quad (2)$$

where Z enforces normalization. For an efficient decimation, we alternate between detection directions nearly coinciding with the vectors associated with q non-orthogonal $|+_p\rangle$ states. Each atom has a chance to reduce the probability of a photon number different from the one decimated by its predecessor. After a finite number of steps, numerical simulations predict that a single n value (modulo $2q$) survives.

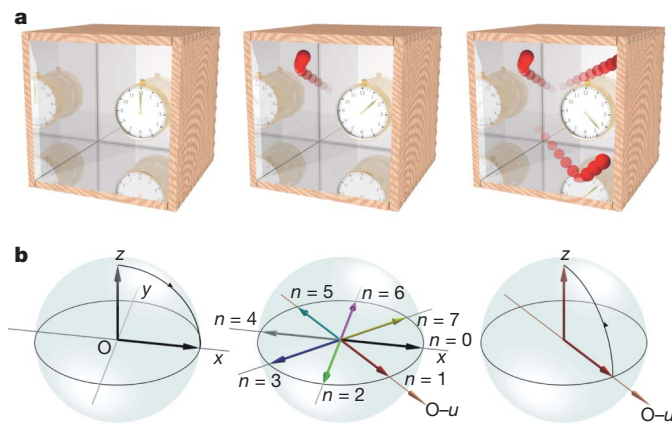


Figure 1 | Principle of QND photon counting. **a**, Thought experiment with a clock in a box containing n photons. The hand of the clock undergoes a $\pi/4$ phase-advance per photon ($n = 0, 1, 3$ represented). **b**, Evolution of the atomic spin on the Bloch sphere in a real experiment: an initial pulse R_1 rotates the spin from O–z to O–x (left). Light shift produces a $\pi/4$ phase shift per photon of the spin's precession in the equatorial plane. Directions associated with $n = 0$ to 7 end up regularly distributed over 360° (centre). Pulse R_2 maps the direction O–u onto O–z, before the atomic state is read out (right).

Observing the field-state collapse

We have applied this procedure to a coherent microwave field at 51.1 GHz stored in an ultrahigh- Q Fabry–Pérot cavity made of niobium-coated superconducting mirrors²⁸. Our set-up is described in ref. 29. The cavity has a very long damping time $T_c = 0.130$ s. It is cooled to 0.8 K (average thermal photon number $n_t = 0.05$). The field is prepared by coupling a short microwave pulse into C (by way of diffraction on the mirrors' edges²⁸). Its photon number distribution and average photon number, $n_0 = 3.82 \pm 0.04$, are inferred from the experimental data (see below). Our single-photon-sensitive spin-clocks are circular Rydberg atoms of rubidium. They cross C successively, separated on average by 2.33×10^{-4} s. Parameters are adjusted to realize a $\sim \pi/4$ clock shift per photon (Methods), corresponding to eight positions of the spin on the Bloch sphere (Fig. 1b). This configuration is adapted to count photon numbers between 0 and 7. For $n_0 = 3.82$, the probability for $n \geq 8$ is 3.5%.

Four phases ϕ_i ($i = a, b, c, d$), corresponding to directions pointing approximately along the spin states associated with $n = 6, 7, 0, 1$, are used, in random order, for successive atoms (Methods). A sequence of j values can be decoded only when combined with the corresponding phase choices, in analogy with the detection basis reconciliation of quantum key distribution protocols³⁶. Figure 2a shows the data from the first 50 detected atoms, presented as (j, i) doublets, for two independent detection sequences performed on the same initial field.

From these real data, we compute the products of functions $\Pi_N(n) = \prod_{(k=1 \dots N)} [A + B \cos(n\Phi - \phi_{i(k)} + j(k)\pi)]$. The A , B , Φ and ϕ_i values are given by Ramsey interferometer calibration

(Methods). The evolutions of $\Pi_N(n)$, displayed as functions of n treated as a continuous variable, are shown in Fig. 2b for N increasing from 1 to 50. The $\Pi_N(n)$ functions converge into narrow distributions whose widths decrease as more information is acquired. These functions are determined uniquely by the experimental data. Their evolution is independent of any a priori knowledge of the initial photon distribution. The data sequence itself, however, depends of the unknown state of the field, which the measurement reveals.

Inserting $\Pi_N(n)$ into equation (2) and extending the procedure to $N = 110$, we obtain the evolution of the photon number histograms for these two realizations (Fig. 2c). These histograms show how our knowledge of the field state evolves in a single measuring sequence, as inferred from bayesian logic. The initial distributions ($P_0(n) = 1/8$) are flat because the only knowledge assumed at the beginning of each sequence is the maximum photon number n_{\max} . Data are analysed after the experiment, but $P_N(n)$ could also be obtained in real time. The progressive collapse of the field into a Fock state (here $|n = 5\rangle$ or $|n = 7\rangle$) is clearly visible. Information extracted from the first 20 to 30 atoms leaves an ambiguity between two competing Fock states. After ~ 50 atoms (detected within ~ 0.012 s), each distribution has turned into a main peak with a small satellite, which becomes totally negligible at the end of the two sequences.

Reconstructing photon number statistics

Repeatedly preparing the field in the same coherent state, we have analysed 2,000 independent sequences, each made of 110 (j, i) doublets recorded within $T_m \approx 0.026$ s. This measuring time is a compromise.

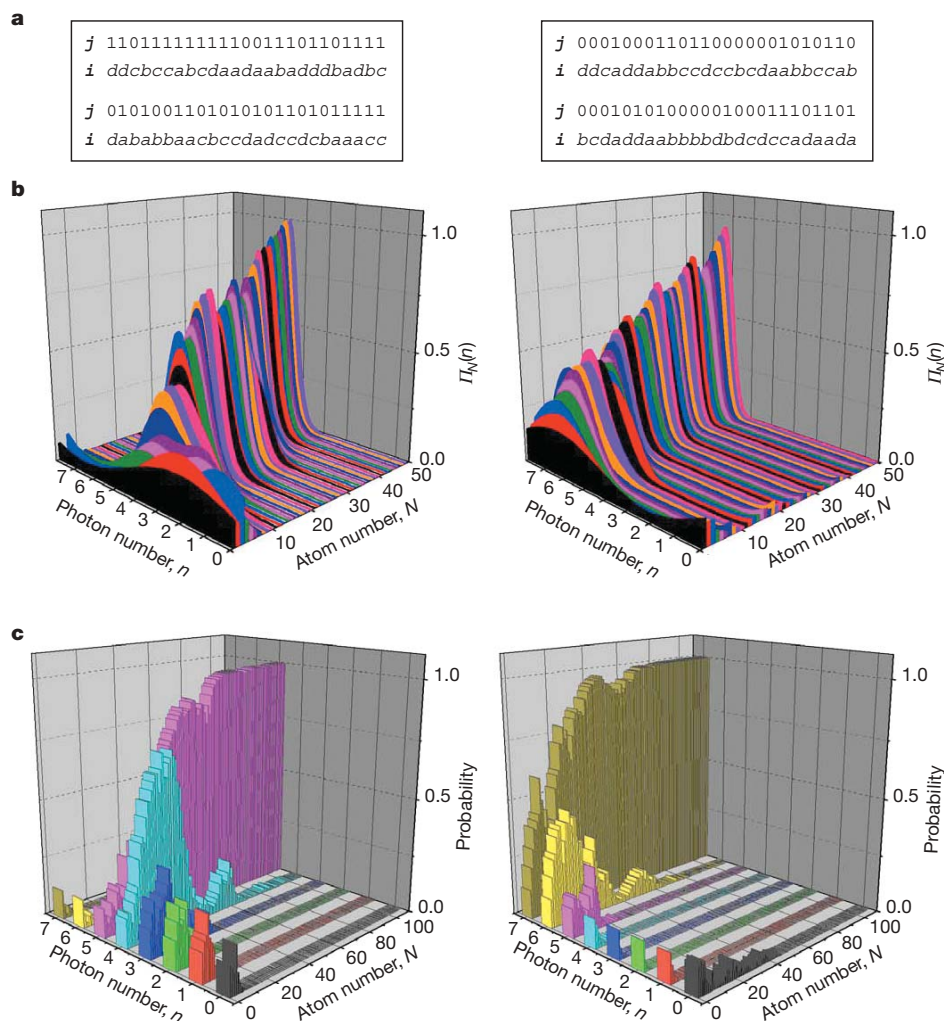


Figure 2 | Progressive collapse of field into photon number state.

a, Sequences of (j, i) data (first 50 atoms) produced by two independent measurements. **b**, Evolution of $\Pi_N(n)$ for the two sequences displayed in **a**, when N increases from 1 to 50, n being treated as a continuous variable (integral of $\Pi_N(n)$ normalized to unity). **c**, Photon number probabilities plotted versus photon and atom numbers n and N . The histograms evolve, as N increases from 0 to 110, from a flat distribution into $n = 5$ and $n = 7$ peaks.

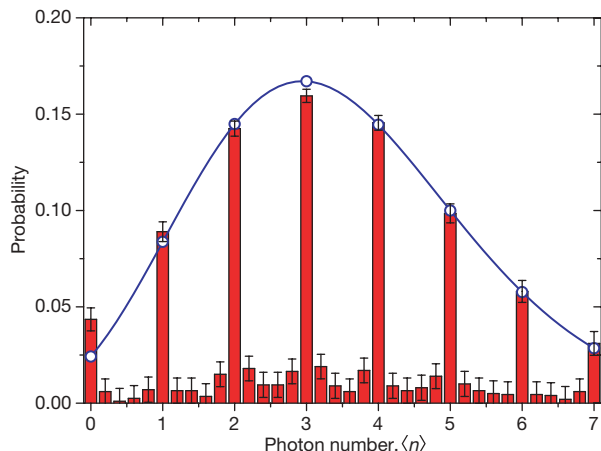


Figure 3 | Reconstructed photon number distribution. Histogram of $\langle n \rangle$ values obtained from 2,000 QND collapse sequences (each involving $N = 110$ atoms). The $\langle n \rangle$ s are sampled in intervals of 0.2. The error bars are the statistical standard deviations. The peaks at integer numbers reveal Fock states. The background is due to incomplete or interrupted collapses. Data shown as blue circles are obtained by fitting the distribution of integer number peaks to a Poisson law, yielding $\langle n \rangle_{\text{ave}} = 3.46 \pm 0.04$ (the blue line represents a continuous Poisson distribution joining the circles as a guide for the eye).

Short compared to T_c , it is long enough to allow for a good convergence of the photon number distribution. We have computed the mean photon number $\langle n \rangle = \sum_n n P_N(n)$ at the end of each sequence. The histogram of these $\langle n \rangle$ values, sampled in bins of width 0.2, is shown in Fig. 3. Peaks at integers appear on top of a small background due to sequences that have not fully collapsed, or that have been interrupted by field decay.

The histogram of the peaks in Fig. 3 directly reveals the photon number probability distribution of the initial field, modified by damping during the measurement. Disregarding the 23% background, we fit the experimental histogram of integer values to a Poisson law with $\langle n \rangle_{\text{ave}} = 3.46 \pm 0.04$ (blue circles), and normalized to 0.77 (probability of fully converged sequences). This is the expected distribution for a coherent field with an initial mean $n_0 = 3.82 \pm 0.04$, after decay during the time $T_m/2 \approx T_c/10$. Remarkably, the non-converged sequences do not introduce any noticeable bias in the distribution of fully collapsed measurements. The experimental excess probability of 0.019 ± 0.006 for $n = 0$ is well understood. It is due to the measurement being performed modulo 8, which attributes $n = 8$ events (0.012 probability) to the $n = 0$ bin. The near-perfect agreement of the fit with the experiment provides a direct verification of the quantum postulate about the probabilities of measurement outcomes.

Repeated measurements and field jumps

Repeatability is another fundamental feature of an ideal QND measurement. To test it, we follow the evolution of the field state along sequences made of $\sim 2,900$ atoms. We determine $P_N(n)$ and $\langle n \rangle$ up to $N = 110$. We then drop the first atom and replace it with the 111st one, resuming the calculation with a flat initial distribution and obtain a new $\langle n \rangle$. We repeat the procedure atom by atom. We thus decode continuously a single field history versus time. Measurements separated by more than T_m exploit independent information.

Figure 4a shows the evolution of $\langle n \rangle$ over 0.7 s for the two sequences whose initial data are displayed in Fig. 2a. In each case, $\langle n \rangle$ evolves quickly towards an integer (5 or 7). This collapse is followed by a plateau, corresponding in these two realizations to ~ 2 independent measurements. Eventually, cavity damping results in a photon loss: a quantum jump occurs, decreasing $\langle n \rangle$ by one. This event is recorded after a delay of a fraction of T_m with respect to the real jump time, as several atoms are required after the quantum leap to build up the new photon number probability. Note in the inset of Fig. 4a that it takes about 0.01 s for the atoms to 'realize' that a jump has occurred. The staircase-like evolution of the field proceeds in this way down to vacuum. Figure 4b presents four other examples of signals following a collapse into $n = 4$. The randomness of the step durations is typical of quantum dynamics. In one of these recordings (leftmost panel), the $n = 4$ step lasts 0.235 s, corresponding to ~ 9 independent QND measurements

The lifetime of a n -Fock state, T_c/n at $T = 0$ K (ref. 39), is reduced by thermal effects to $T_c/[n + n_t(2n + 1)]$, that is, ~ 0.029 s, ~ 0.023 s and ~ 0.017 s for $n = 4, 5$ and 7, respectively. The statistical analysis of 2,000 QND sequences, each of 0.7 s duration, provides a detailed description of the dynamical evolution under cavity relaxation of Fock states with n up to 7 (J.B. *et al.*, manuscript in preparation). The sequences of Fig. 4, in which some Fock states survive much longer than their lifetimes, are relatively rare events. We have selected them to demonstrate the ability of the QND procedure to generate and repeatedly measure large n -Fock states of radiation in a cavity.

Beyond energy measurements

Our QND source of Fock states operates in a way different from previous methods based on resonant^{25,40} or Raman⁴¹ processes in cavity quantum electrodynamics, which have so far been limited to smaller photon numbers ($n = 1$ or 2). This QND measurement opens novel perspectives for the generation of non-classical states of light. If the initial photon number distribution spans a range of n values larger than $2q$, the decimations induced by successive atoms do not distinguish between n and $n + 2q$. The field then collapses into a superposition of the form $\sum_q c_{n+2q} |n + 2q\rangle$. For instance, $c_0 |0\rangle + c_{2q} |2q\rangle$

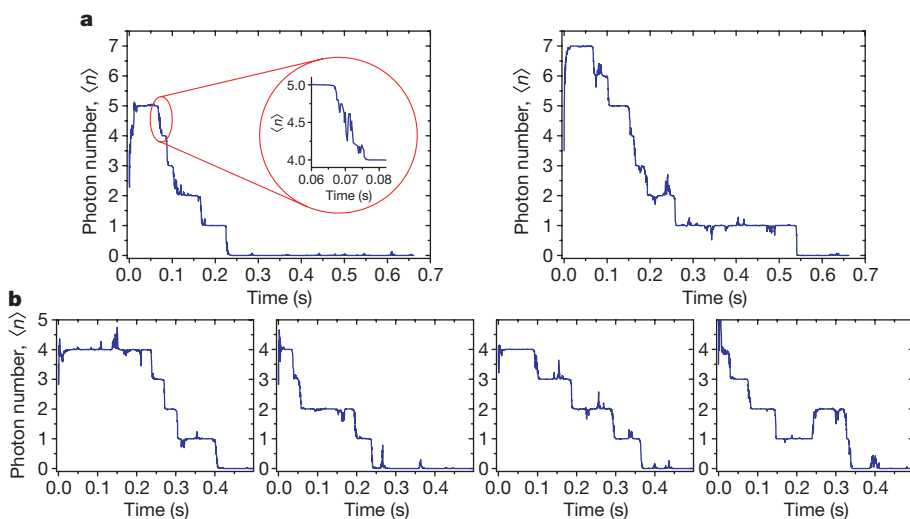


Figure 4 | Repeated QND measurements.

a, Mean photon number $\langle n \rangle$ followed over 0.7 s for the two sequences whose collapse is analysed in Fig. 2. After converging, $\langle n \rangle$ remains steady for a while, before successive quantum jumps bring it down to vacuum. Inset, zoom into the $n = 5$ to 4 jump, showing that it is detected in a time of ~ 0.01 s. **b**, Four other signals recording the evolution of $\langle n \rangle$ after field collapse into $n = 4$. Note in the leftmost frame the exceptionally long-lived $n = 4$ state, and in the rightmost frame the $n = 1$ to 2 jump revealing a thermal field fluctuation.

represents a field coherently suspended between vacuum and $2q$ photons. This superposition of states with energies differing by many quanta is a new kind of 'Schrödinger cat' state of light.

Superpositions of field states with the same amplitude but different phases—'Schrödinger cats' of a different kind—are also generated in this experiment. As the photon number converges, its conjugate variable, the field's phase, gets blurred. After the first atom's detection, the initial state collapses into a superposition of two coherent states with different phases^{31,42}. Each of these components is again split into two coherent states by the next atom and so on, leading to complete phase-uncertainty when the photon number has converged³¹. The evolution of the Schrödinger cat states generated in the first steps of this process could be studied by measuring the field Wigner function⁴³. Decoherence^{44,45} of superpositions of coherent states containing many photons could be monitored in this way.

METHODS SUMMARY

The preparation and detection of circular Rydberg atoms, the cavity and the Ramsey interferometer are described elsewhere^{20,26,29}. The $|\pm\rangle$ states are the circular Rydberg levels of rubidium with principal quantum numbers 51 and 50 (transition frequency ~ 51.1 GHz). The theoretical phase shift per photon^{20,26} is $\Omega^2 t / \delta$, where $\Omega / 2\pi = 50$ kHz is the vacuum Rabi frequency at cavity centre, $\delta / 2\pi$ is the atom–cavity detuning and $t = 3 \times 10^{-5}$ s is the effective atom–cavity interaction time. It is defined as $t = (\pi/2)^{1/2} w / v$, where w is the waist of the gaussian cavity field mode ($w = 6$ mm) and v the atomic velocity ($v = 250$ m s⁻¹). This effective time is obtained by averaging the spatial variation of the square of the atom–field coupling as the atom crosses the cavity mode^{20,26}.

More details about the experimental settings, including the determination of the A and B parameters, the fine tuning of the phase shift per photon, Φ , and the adjustment of the four phases of the Ramsey interferometer are given in the Methods section. We also analyse the adiabaticity of the atom–field coupling, which is an essential feature of our measurement. We describe the generation of sequences of atoms crossing the cavity one at a time with a well-defined velocity, and we discuss the effect of rare multi-atom events on the detection signals. We also explain why the sequences of detection directions ϕ_i occur randomly in a measuring sequence. We conclude by discussing alternative strategies to pin down the photon number non-destructively.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 May; accepted 28 June 2007.

- Wheeler, J. A. & Zurek, W. H. *Quantum Theory and Measurement* (Princeton Series in Physics, Princeton, New Jersey, 1983).
- Kuzmich, A., Mandel, L. & Bigelow, N. P. Generation of spin squeezing via continuous quantum non-demolition measurement. *Phys. Rev. Lett.* **85**, 1594–1597 (2000).
- Geremia, J. M., Stockton, J. K. & Mabuchi, H. Suppression of spin projection noise in broadband atomic magnetometry. *Phys. Rev. Lett.* **94**, 203002 (2005).
- Braginsky, V. B. & Vorontsov, Y. I. Quantum mechanical limitations in macroscopic experiments and modern experimental technique. *Usp. Fiz. Nauk* **114**, 41–53 (1974); *Sov. Phys. Usp.* **17**, 644–650 (1975).
- Thorne, K. S., Drever, R. W. P., Caves, C. M., Zimmerman, M. & Sandberg, V. D. Quantum nondemolition measurements of harmonic oscillators. *Phys. Rev. Lett.* **40**, 667–671 (1978).
- Unruh, W. G. Analysis of quantum nondemolition measurement. *Phys. Rev. D* **18**, 1764–1772 (1978).
- Caves, C. M., Thorne, K. S., Drever, R. W. P., Sandberg, V. D. & Zimmerman, M. On the measurement of a weak classical force coupled to a quantum mechanical oscillator. *Rev. Mod. Phys.* **52**, 341–392 (1980).
- Milburn, G. J. & Walls, D. F. Quantum nondemolition measurements via quadratic coupling. *Phys. Rev. A* **28**, 2065–2070 (1983).
- Peil, S. & Gabrielse, G. Observing the quantum limit of an electron cyclotron: QND measurement of quantum jumps between Fock states. *Phys. Rev. Lett.* **83**, 1287–1290 (1999).
- Leibfried, D., Blatt, R., Monroe, C. & Wineland, D. J. Quantum dynamics of single trapped ions. *Rev. Mod. Phys.* **75**, 281–324 (2003).
- Schmidt, P. O. *et al.* Spectroscopy using quantum logic. *Science* **309**, 749–752 (2005).
- Kuzmich, A. *et al.* Quantum non-demolition measurements of collective atomic spin. *Phys. Rev. A* **60**, 2346–2350 (1999).
- Grangier, P., Levenson, J. A. & Poizat, J. P. Quantum non-demolition measurements in optics. *Nature* **396**, 537–542 (1998).
- Imoto, N., Haus, H. A. & Yamamoto, Y. Quantum nondemolition measurement of the photon number via the optical Kerr effect. *Phys. Rev. A* **32**, 2287–2292 (1985).
- Grangier, P., Roch, J. F. & Roger, G. Observation of backaction-evading measurement of an optical intensity in a three-level atomic non-linear system. *Phys. Rev. Lett.* **66**, 1418–1421 (1991).
- Roch, J. F. *et al.* Quantum nondemolition measurements using cold trapped atoms. *Phys. Rev. Lett.* **78**, 634–637 (1997).
- Bencheikh, K., Levenson, J. A., Grangier, P. & Lopez, O. Quantum nondemolition demonstration via repeated backaction evading measurements. *Phys. Rev. Lett.* **75**, 3422–3425 (1995).
- Bruckmeier, R., Schneider, K., Schiller, S. & Mlynek, J. Quantum nondemolition measurements improved by a squeezed meter input. *Phys. Rev. Lett.* **78**, 1243–1246 (1997).
- Haroche, S. in *Fundamental Systems in Quantum Optics* (eds Dalibard, J. & Raimond, J. M.) 767–940 (Les Houches Summer School, Session LIII, North Holland, Amsterdam, 1992).
- Haroche, S. & Raimond, J. M. *Exploring the Quantum: Atoms, Cavities and Photons* (Oxford Univ. Press, Oxford, UK, 2006).
- Walther, H., Varcoe, B. T. H., Englert, B. G. & Becker, T. Cavity quantum electrodynamics. *Rep. Prog. Phys.* **69**, 1325–1382 (2006).
- Birnbaum, K. M. *et al.* Photon blockade in an optical cavity with one trapped atom. *Nature* **436**, 87–90 (2005).
- Wilk, T., Webster, S. C., Kuhn, A. & Rempe, G. Single-atom single-photon quantum interface. *Science* doi:10.1126/science.1143835 (published online 21 June 2007).
- Nogues, G. *et al.* Seeing a single photon without destroying it. *Nature* **400**, 239–242 (1999).
- Varcoe, B. T. H., Brattke, S., Weidinger, M. & Walther, H. Preparing pure photon number states of the radiation field. *Nature* **403**, 743–746 (2000).
- Raimond, J. M., Brune, M. & Haroche, S. Manipulating quantum entanglement with atoms and photons in a cavity. *Rev. Mod. Phys.* **73**, 565–582 (2001).
- Schuster, D. I. *et al.* Resolving photon number states in a superconducting circuit. *Nature* **445**, 515–518 (2007).
- Kuhr, S. *et al.* Ultrahigh finesse Fabry–Pérot superconducting resonator. *Appl. Phys. Lett.* **90**, 164101 (2007).
- Gleyzes, S. *et al.* Quantum jumps of light recording the birth and death of a photon in a cavity. *Nature* **446**, 297–300 (2007).
- Brune, M., Haroche, S., Lefèvre, V., Raimond, J. M. & Zagury, N. Quantum non-demolition measurement of small photon numbers by Rydberg atom phase-sensitive detection. *Phys. Rev. Lett.* **65**, 976–979 (1990).
- Brune, M., Haroche, S., Raimond, J. M., Davidovich, L. & Zagury, N. Quantum non-demolition measurements and generation of Schrödinger cat states. *Phys. Rev. A* **45**, 5193–5214 (1992).
- Carmichael, H. *An Open System Approach to Quantum Optics* (Springer, Berlin, 1993).
- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, UK, 2000).
- Glauber, R. J. Coherent and incoherent states of the radiation field. *Phys. Rev.* **131**, 2766–2788 (1963).
- D'Ariano, G. M. & Yuen, H. P. Impossibility of measuring the wave function of a single quantum system. *Phys. Rev. Lett.* **76**, 2832–2835 (1996).
- Gisin, N., Ribordy, G. G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74**, 145–195 (2002).
- Ramsey, N. *Molecular Beams* (Oxford Univ. Press, Oxford, UK, 1984).
- Bayes, T. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. R. Soc. Lond.* **53**, 370–418 (1763).
- Lu, N. Effects of dissipation on photon statistics and the lifetime of a pure number state. *Phys. Rev. A* **40**, 1707–1708 (1989).
- Maitre, X. *et al.* Quantum memory with a single photon in a cavity. *Phys. Rev. Lett.* **79**, 769–772 (1997).
- Bertet, P. *et al.* Generating and probing a two-photon Fock state with a single atom in a cavity. *Phys. Rev. Lett.* **88**, 143601 (2002).
- Brune, M. *et al.* Observing the progressive decoherence of the meter in a quantum measurement. *Phys. Rev. Lett.* **77**, 4887–4890 (1996).
- Bertet, P. *et al.* Direct measurement of the Wigner function of a one photon Fock state in a cavity. *Phys. Rev. Lett.* **89**, 200402 (2002).
- Giulini, D. *et al.* *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer, Berlin, 1996).
- Zurek, W. H. Decoherence, einselection and the quantum origins of the classical. *Rev. Mod. Phys.* **75**, 715–775 (2003).

Acknowledgements This work was supported by the Agence Nationale pour la Recherche (ANR), by the Japan Science and Technology Agency (JST), and by the EU under the IP projects SCALA and CONQUEST. C.G. and S.D. were funded by the Délégation Générale à l'Armement (DGA). J.-M.R. is a member of the Institut Universitaire de France (IUF).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.B. (brune@lkb.ens.fr) or S.H. (haroche@lkb.ens.fr).

METHODS

Experimental settings. To calibrate the Ramsey interferometer, we send atoms prepared in $|-z\rangle$ across the set-up, with C empty. We measure the probability $P(j=0, \phi)$ for detecting the atom in $|+z\rangle$ as a function of the relative phase ϕ between R_1 and R_2 . From the Ramsey fringes, we obtain the phase origin $\phi = 0$ corresponding to a detection along the $|+z\rangle$ spin direction.

Tuning of δ is performed by moving the cavity mirrors with piezoelectric actuators. In theory, a $\pi/4$ phase shift corresponds to $\delta/2\pi = 300$ kHz. We set the detuning close to this value (with a 15 kHz uncertainty due to imperfect knowledge of the atomic transition frequency affected by small residual Stark and Zeeman shifts). With this detuning, the photon-induced phase shift is, within 3%, a linear function of n for $n = 0$ to 7.

By Stark-shifting the atomic spin phase with a short electric field pulse applied just before R_2 and adjusted to different amplitudes, we translate the fringes and set ϕ to four different values, close to $l\pi/4$ ($l = -2, -1, 0$ and 1). After preliminary settings of Φ and ϕ_b we refine our calibration. We inject a small coherent field in C ($n_0 = 1.2$). This field has a negligible probability for $n > 4$. For each ϕ_b we record the fraction $\eta_0(\phi_i)$ of spins found in $|+z\rangle$ on a sequence of atoms crossing C in a time short compared to T_c . Repeating the sequence many times, we find distributions of $\eta_0(\phi_i)$ s, which we fit as a sum of five peaks centred on the discrete values equal to $P^{(\text{exp})}(j=0, \phi_i/n) = [A + B\cos(n\Phi - \phi_i)]/2$, with $n = 0$ to 4. From a best fit of the η_0 distributions, we get the values $A = 0.907 \pm 0.004$, $B = 0.674 \pm 0.004$, $\Phi/\pi = 0.233 \pm 0.004$, $\phi_i/\pi = -0.464 \pm 0.013$, -0.229 ± 0.009 , -0.015 ± 0.007 and $+0.261 \pm 0.006$ ($i = a, b, c, d$). These values are inserted in equation (2). The fringe contrast B is reduced below 1 by experimental imperfections (stray fields, detection errors, two-atom events, see below). The other six parameters A , Φ/π and ϕ_i/π ($i = a, b, c, d$) are close to their ideal values ($1, 1/4, -1/2, -1/4, 0, 1/4$, respectively).

Adiabaticity. The adiabatic variation of the coupling as the atoms cross the gaussian profile of the cavity mode keeps the atomic emission rate extremely low. The theoretical probability that an atom deposits an additional photon when the cavity contains a coherent field with $n_0 = 3.82$ is below 1.3×10^{-6} . Consistent with this very small value, we found on analysing our experimental data that the average number of photons deposited in C by a sequence of $\sim 2,900$ atoms is negligible compared to n_i .

Atomic sequences. They are realized by pulsing the Rydberg atom preparation. The atoms are excited from a thermal atomic beam, velocity selected by optical pumping, at a rate of 1.4×10^4 pulses per second. The velocity spread, $\Delta v = \pm 1 \text{ m s}^{-1}$ around 250 m s^{-1} , has a negligible effect on the Ramsey fringe contrast. In order to limit the number of events with two atoms per pulse, the intensity of the exciting lasers is kept low (average number of detected Rydberg atoms per pulse is 0.3, detection efficiency 50%). Undetected atoms do not affect the photon number distribution. A single atom is counted in 22% of the preparation pulses, while 3% of them contain a detected atom pair. When two atoms (whether detected or not) cross C together there is a slight reduction of the interferometer contrast, owing to small cavity-mediated interactions⁴⁶ between the atoms. This reduction is taken into account in the measured B value. All detected events with one or two atoms per pulse are compiled independently in the data analysis. When three atoms are in C together the fringe contrast is strongly reduced, but the probability of these events is small (2% probability for preparing 3 atoms or more per pulse).

Randomness of detection directions. The interferometer phase is changed from pulse to pulse, going cyclically from ϕ_a to ϕ_d . As the presence of one (or two) atoms in a given pulse is random, we cannot predict which phase will correspond to the next observed atom. We acquire this knowledge by detecting the atom, and $i(k)$ is thus a randomly measured variable.

Other QND measurement strategies. Efficient photon number decimation could be obtained by alternative methods. As suggested in refs 30 and 31, we could change the atom–cavity interaction time (and hence the phase shift per photon) by detecting randomly atoms from a thermal atomic beam, without velocity selection. The optimal data acquisition procedure consists in applying to successive atoms a sequence of $\pi, \pi/2, \pi/4, \dots$ phase shifts per photon, while adjusting ϕ for each spin, based on the result of the previous measurement. This expresses n in binary code, each atom providing a bit of information^{20,47}. The required number of atoms per measuring sequence is then minimal, equal to the smallest integer $\geq \log_2(n_{\text{max}} + 1)$. This ideal strategy requires however a deterministic beam of atoms, with perfect Ramsey fringe contrast and 100% detection efficiency.

46. Osnaghi, S. *et al.* Coherent control of an atomic collision in a cavity. *Phys. Rev. Lett.* **87**, 037902 (2001).

47. Haroche, S., Brune, M. & Raimond, J. M. Measuring photon numbers in a cavity by atomic interferometry: optimizing the convergence procedure. *J. Phys. II France* **2**, 659–670 (1992).

ARTICLES

Cortico-striatal synaptic defects and OCD-like behaviours in *Sapap3*-mutant mice

Jeffrey M. Welch^{1*}, Jing Lu^{1,4*}, Ramona M. Rodriguiz², Nicholas C. Trotta¹, Joao Peca^{1,5}, Jin-Dong Ding⁶, Catia Feliciano^{1,7}, Meng Chen³, J. Paige Adams⁸, Jianhong Luo⁴, Serena M. Dudek⁸, Richard J. Weinberg⁶, Nicole Calakos^{1,3}, William C. Wetsel² & Guoping Feng¹

Obsessive-compulsive disorder (OCD) is an anxiety-spectrum disorder characterized by persistent intrusive thoughts (obsessions) and repetitive actions (compulsions). Dysfunction of cortico-striato-thalamo-cortical circuitry is implicated in OCD, although the underlying pathogenic mechanisms are unknown. SAP90/PSD95-associated protein 3 (SAPAP3; also known as DLGAP3) is a postsynaptic scaffolding protein at excitatory synapses that is highly expressed in the striatum. Here we show that mice with genetic deletion of *Sapap3* exhibit increased anxiety and compulsive grooming behaviour leading to facial hair loss and skin lesions; both behaviours are alleviated by a selective serotonin reuptake inhibitor. Electrophysiological, structural and biochemical studies of *Sapap3*-mutant mice reveal defects in cortico-striatal synapses. Furthermore, lentiviral-mediated selective expression of *Sapap3* in the striatum rescues the synaptic and behavioural defects of *Sapap3*-mutant mice. These findings demonstrate a critical role for SAPAP3 at cortico-striatal synapses and emphasize the importance of cortico-striatal circuitry in OCD-like behaviours.

OCD, a common and incapacitating psychiatric disorder, affects ~2 per cent of the world population^{1,2}. OCD is characterized by persistent intrusive thoughts (obsessions), repetitive actions (compulsions) and excessive anxiety. Clinical expression of OCD is heterogeneous in the types of obsessions and compulsions, heritability and co-morbid conditions, probably reflecting heterogeneity in the underlying pathology³. In addition, there are many disorders that share features with OCD, termed 'OC-spectrum disorders', which include Tourette's syndrome, trichotillomania and body dysmorphic disorder.

The neurobiological basis of OCD is unclear. However, lesions, functional neuro-imaging, and neuropsychological studies have indicated that the cortico-striato-thalamo-cortical circuitry may have a key role in the pathogenesis of OCD^{3–5}. Familial studies of OCD indicate that the risk to first degree relatives is 3–12 times greater than the general population, similar to familial risk rates observed for bipolar disorder and schizophrenia^{6–8}. In addition, concordance for OCD is greater among pairs of monozygotic (80–87%) than dizygotic (47–50%) twins^{9,10}. However, no genetic factors have yet been identified as a cause of OCD, presumably reflecting aetiological heterogeneity within the disorder. Although many of the symptoms of OCD can be ameliorated by enhancing serotonin neurotransmission, it is not clear at the cellular level whether defects in the serotonergic system are the primary cause^{3–5}. Indeed, dopaminergic and glutamatergic neurotransmitter systems have also been implicated^{5,11}. Here we show that targeted deletion of *Sapap3* in mice leads to a behavioural phenotype similar to OCD: compulsive overgrooming behaviour, increased anxiety, and response to selective serotonin reuptake inhibitors. In these mutant mice, we have identified defects at cortico-striatal synapses, part of the circuitry implicated in OCD, and we show that selective expression of *Sapap3* in the

striatum rescues the synaptic and behavioural defects. These findings indicate that defects in excitatory transmission at cortico-striatal synapses may underlie some aspects of OCD.

Sapap3^{−/−} mice exhibit self-inflicted facial lesions

SAPAP family proteins were originally identified as postsynaptic density (PSD) components that interact with the PSD95 and Shank families of proteins^{12,13}, two other multi-domain postsynaptic scaffolding proteins at excitatory synapses. Together, these three groups of proteins are thought to form a key scaffolding complex that regulates the trafficking and targeting of neurotransmitter receptors and signalling molecules to the postsynaptic membrane of excitatory synapses^{14–16}. There are four highly homologous genes encoding members of the SAPAP family¹³. Of these, *Sapap1*, 3 and 4 are highly, but differentially, expressed in several regions of the brain^{17,18}. Notably, *Sapap3* is the only member highly expressed in the striatum. To facilitate study of the *in vivo* function of SAPAP proteins at synapses, we generated knockout mice for the *Sapap3* gene using homologous recombination in mouse embryonic stem cells (Supplementary Fig. 1a–d). Mice homozygous for the *Sapap3* deletion (*Sapap3*^{−/−}) were born at the expected mendelian rate, grew to adulthood with body weights similar to wild-type mice, and were fertile. Anatomical and histological analyses of brain showed that *Sapap3*^{−/−} mice were grossly normal (data not shown).

By the age of 4–6 months, however, *Sapap3*^{−/−} mice developed lesions on their head, neck and snout regions (Fig. 1a). This phenotype was 100% penetrant. Lesions were usually first noticed as a patch of hairless skin under the eyes, or swelling of the snout, progressing to relatively symmetric bilateral lesions encompassing large parts of the neck and head. *Sapap3*^{−/−} mice developed lesions regardless of

¹Department of Neurobiology, ²Departments of Psychiatry and Behavioral Sciences, Cell Biology, Neurobiology, and Mouse Behavioral and Neuroendocrine Analysis Core Facility, ³Division of Neurology, Center for Translational Neuroscience, Duke University Medical Center, Durham, North Carolina 27710, USA. ⁴Department of Neurobiology, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310058, China. ⁵Center for Neuroscience and Cell Biology, University of Coimbra, 3004-517 Coimbra, Portugal. ⁶Department of Cell and Developmental Biology, Neuroscience Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA. ⁷Gulbenkian PhD Programme in Biomedicine, Gulbenkian Science Institute, 2781-901 Oeiras, Portugal. ⁸National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA. *These authors contributed equally to this work.

whether they were housed alone or with cage mates. Lesions were not observed on wild-type or heterozygous mice even when housed in the same cage with *Sapap3*^{-/-} mice from birth, thereby excluding the possibility that lesions were caused by allogrooming or were the result of aggressive encounters with other *Sapap3*^{-/-} mice. In fact, *Sapap3*^{-/-} mice were not observed to behave aggressively, but were often seen engaged in self-grooming whether they were housed alone or in groups.

Given the obvious lesions on the *Sapap3*^{-/-} mice, we tested the possibility that their lesions might be caused by peripheral cutaneous defects, such as inflammation or abnormal afferent sensation. We examined facial skin from pre-lesion *Sapap3*^{-/-} mice that exhibited increased grooming. Histological analysis of skin did not reveal any anatomical differences among wild-type, *Sapap3*^{+/-} and *Sapap3*^{-/-} mice, and no lymphocytic or granulocytic infiltration was observed

(Supplementary Fig. 1e, f). We did, however, find lymphocytic/granulocytic infiltration in skin with lesions, probably owing to injury and infection (Supplementary Fig. 1g). No differences were detected in sensory innervation among wild-type, *Sapap3*^{+/-} and *Sapap3*^{-/-} mice; hair-nerve end organs, lobular corpuscle-like nerve endings, and free epidermal nerve endings were present and not different between *Sapap3*^{-/-} and control mice (Supplementary Fig. 1h, i). Thus, our examination revealed no obvious peripheral defects that would indicate a cause for the lesions. Together, these findings raised the possibility that *Sapap3*^{-/-} mice have excessive and injurious self-grooming behaviour.

OCD-like behaviours of *Sapap3*^{-/-} mice

To determine whether *Sapap3*^{-/-} mice groom excessively, we continuously videotaped habituated, individually housed mice for 24 h. Isolated *Sapap3*^{-/-} mice showed dramatically increased grooming bouts and spent significantly more time self-grooming than wild-type litter mates (Fig. 1b, c). The increased grooming was present throughout the day, including the period when mice usually sleep (10:00–14:00). Both mutant mice with facial lesions and mice yet to develop lesions showed similar degrees of increased grooming (Fig. 1d), indicating that the lesion itself was not the cause of increased grooming. We thus conclude that *Sapap3*^{-/-} mice have excessive and injurious levels of self-grooming, a phenotype reminiscent of compulsive behaviours.

Because the *Sapap3*^{-/-} mice exhibited a compulsive-like behaviour, we considered whether their phenotypes further resemble OCD-spectrum disorders. We therefore examined other OCD-like behaviours that are measurable in mice, such as increased anxiety and response to typical pharmacological agents. To test whether *Sapap3*^{-/-} mice exhibit increased anxiety, we conducted open field, dark–light emergence, and elevated zero maze tests. In the open field, mice with anxiety-like phenotypes tend to stay along the walls, avoiding the centre zone. *Sapap3*^{-/-} mice spent much less time exploring the centre area than wild-type litter mates (Fig. 1e). However, activities along the walls and corners, areas believed to be less stressful, were not different between wild-type and *Sapap3*^{-/-} mice (Fig. 1f). In the dark–light emergence test, the latency to cross from a dark into a brightly lit chamber (a stressful environment), and the time spent in the brightly lit chamber were examined. *Sapap3*^{-/-} mice took longer to cross from the dark to the lit chamber and spent less time in the lit chamber than wild-type controls (Fig. 1g, h), even though total activity in the chambers was similar between wild-type and *Sapap3*^{-/-} mice (Fig. 1i). Finally, in the elevated zero maze test, we found that *Sapap3*^{-/-} mice took longer to cross into the open areas (riskier environment) and spent significantly less time exploring the open areas compared with wild-type controls (Fig. 1j, k), whereas total activity was not different between the two genotypes (Fig. 1l). Together, these results indicate that *Sapap3*^{-/-} mice have an anxiety-like phenotype.

We next evaluated whether drugs used to treat OCD would be effective in reducing the abnormal behaviour in *Sapap3*^{-/-} mice. Selective serotonin reuptake inhibitors are a first-line treatment for OCD; although their mechanism of action is not well understood, these drugs alleviate symptoms in approximately 50% of OCD patients³. Fluoxetine treatment for 6 days (once per day at 5 mg kg⁻¹, intraperitoneally) did not affect grooming behaviour of wild-type mice. However, this treatment significantly reduced the excessive grooming in *Sapap3*^{-/-} mice (Fig. 2a). Furthermore, fluoxetine dramatically reduced anxiety-like behaviours in *Sapap3*^{-/-} mice in the dark–light emergence test without affecting their total activity (Fig. 2b, c). In contrast, a single injection of fluoxetine (5 mg kg⁻¹, intraperitoneally) had no effects on their excessive grooming behaviour (Fig. 2d). Thus, *Sapap3*^{-/-} mice exhibit compulsive-like grooming and increased anxiety, both of which are alleviated by 6 days of fluoxetine treatment.

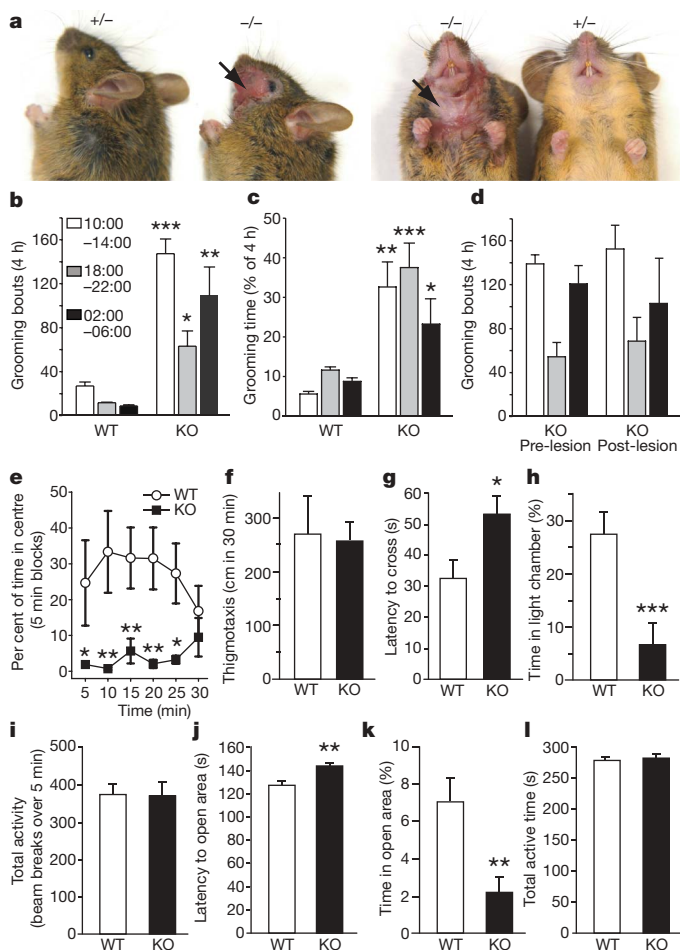


Figure 1 | Facial lesions, excessive grooming and anxiety-like behaviours in *Sapap3*-mutant mice. **a**, *Sapap3*^{-/-} mice have facial and neck skin lesions (arrows). **b**, **c**, *Sapap3*^{-/-} mice (KO) showed more grooming bouts (**b**) and spent more time in self-grooming (**c**) than wild-type mice (WT) at all times examined. **d**, Pre-lesion and post-lesion groups of *Sapap3*^{-/-} mice showed similar degrees of increased grooming. **e**, **f**, In the open field test, *Sapap3*^{-/-} mice spent less time in the centre (**e**), whereas locomotion along the perimeter was comparable to wild-type controls (**f**). **g**–**i**, In the dark–light emergence test, *Sapap3*^{-/-} mice took longer to cross from the dark to the brightly lit chamber (**g**) and spent less time in the brightly lit chamber (**h**), whereas total activity in both chambers was similar (**i**). **j**–**l**, In the elevated zero maze, *Sapap3*^{-/-} mice took longer to cross into the open areas (**j**) and spent less time in these areas than wild-type controls (**k**), whereas total activity in both open and closed areas was similar (**l**). **P* < 0.05, ***P* < 0.01, ****P* < 0.001, repeated measures ANOVA for **b**–**e** and two-tailed *t*-test for **f**–**l**; all data are presented as means ± s.e.m. from 8–9 mice per genotype; Cohen's kappa for intra-observer agreement exceeded 0.92.

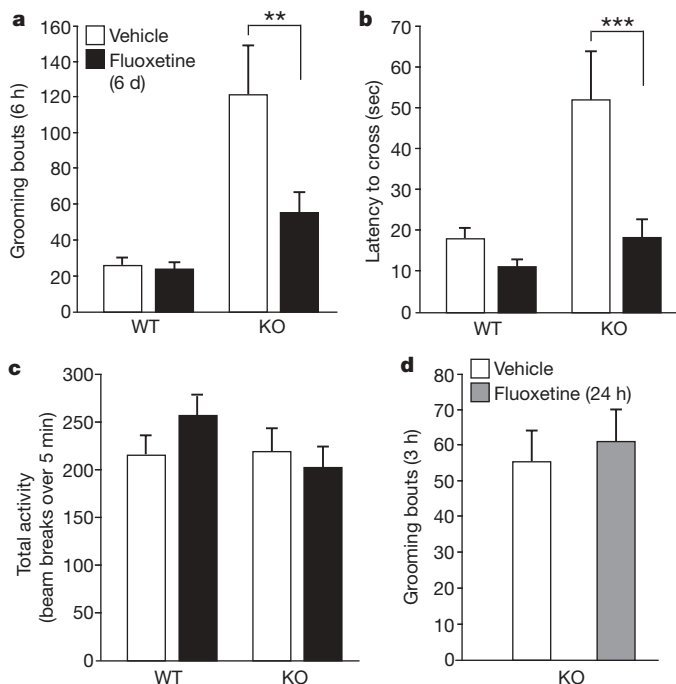


Figure 2 | Fluoxetine treatment alleviates excessive grooming and anxiety-like behaviour. **a**, Daily fluoxetine treatment over six days reduced grooming behaviour in *Sapap3*^{-/-} mice. **b**, **c**, Given over 6 days, fluoxetine alleviated anxiety-like behaviour of *Sapap3*^{-/-} mice in the dark–light emergence test (**b**) without affecting their total activity (**c**). **d**, A single injection of fluoxetine had no effect on the grooming behaviour of *Sapap3*^{-/-} mice. ***P* < 0.01, ****P* < 0.001, repeated measures ANOVA; all data are presented as means ± s.e.m. from 9–11 mice per group; Cohen's kappa for intra-observer agreement exceeded 0.92.

Altered cortico-striatal synaptic transmission

The observations that SAPAP3 is the only member of the SAPAP family highly expressed in the striatum (Fig. 3a) and that *Sapap3*^{-/-} mice show OCD-like behaviours indicate that there

may be defects in striatal neurotransmission in *Sapap3*^{-/-} mice. Because SAPAPs are postsynaptic proteins of excitatory synapses^{12,17} and directly bind PSD95 family proteins, which are known to regulate the trafficking of both AMPA- and NMDA-type glutamate receptors (AMPA and NMDAR, respectively)^{19–21}, we focused on cortico-striatal synapses, which constitute the large majority of glutamatergic synapses in the striatum. We performed extracellular recordings from acute striatal brain slices. A field recording electrode was placed in the dorso-lateral striatum and a stimulating electrode was placed nearby in the corpus callosum. Recordings were obtained in the presence of picrotoxin, a GABA (γ-aminobutyric acid)_A receptor antagonist, to block contaminating responses from intra-striatal GABAergic circuitry. We found that field excitatory postsynaptic potentials (fEPSPs) were significantly reduced in *Sapap3*^{-/-} mice compared with wild-type litter mates (Fig. 3b). Axonal excitability and presynaptic function were not different from wild type, as indicated by the slope of the input–output curves normalized to the peak response, the relationship of stimulation intensity to the amplitude of the action potential component (NP1; Supplementary Fig. 2a, b), and paired-pulse ratios (Fig. 3c), indicating that the reduction in total field responses was probably due to a postsynaptic impairment in synaptic transmission.

We next evaluated NMDAR-dependent responses by recording in the presence of an AMPAR antagonist (NBQX) and the NMDAR co-factor glycine, in the absence of magnesium. Surprisingly, in contrast to the reduction we observed in the total fEPSP amplitude, which is dominated by AMPAR transmission under these conditions, the NMDAR-dependent fEPSPs were elevated in *Sapap3*^{-/-} mice (Fig. 3d, e). These findings indicate a differential effect of *Sapap3* deletion on extracellular field potentials that are dependent on AMPAR and NMDAR activity. We further tested whether the synaptic defects are unique to the striatum by examining synaptic transmission in the hippocampus, in which *Sapap1*, 3 and 4 are highly expressed^{17,18}. We found no defects in CA3–CA1 hippocampal basal synaptic transmission in *Sapap3*^{-/-} mice (Supplementary Fig. 2c, d), probably reflecting functional redundancy or compensation by other members of the SAPAP family. Together, these findings identify an important role for SAPAP3 in postsynaptic glutamatergic synaptic function at cortico-striatal synapses.

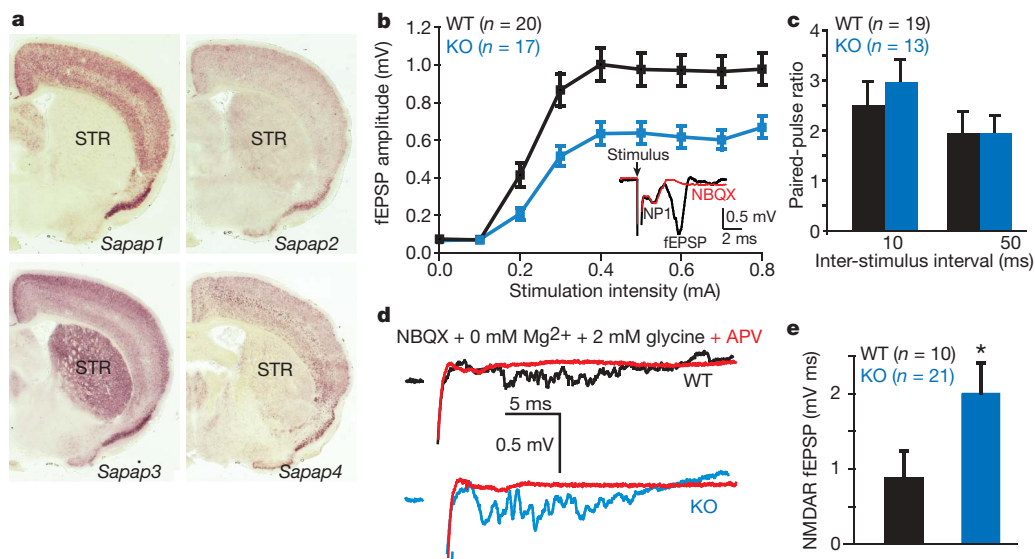


Figure 3 | Altered cortico-striatal synaptic transmission in *Sapap3*-mutant mice. **a**, Only *Sapap3* messenger RNA is highly expressed in the striatum (STR). **b**, Cortico-striatal field potential recordings of acute brain slices show decreased fEPSPs in *Sapap3*^{-/-} mice. *P* < 0.001, repeated measures ANOVA. Inset shows a typical field recording (black trace) and sensitivity to NBQX (red trace). NP1, negative peak 1. **c**, Paired-pulse ratios (slope of

fEPSP no. 2/slope fEPSP no. 1) are similar between wild-type and *Sapap3*^{-/-} mice. **d**, Example traces of NMDAR-dependent fEPSPs recorded in the presence of 50 μM NBQX, 0 mM Mg²⁺ and 2 mM glycine. Sensitivity to APV is indicated by the red trace. **e**, NMDAR fEPSP area is increased in *Sapap3*^{-/-} mice. **P* < 0.05, two-tailed *t*-test; number of recordings in parentheses. All data are presented as means ± s.e.m.

Altered NMDAR composition of the PSD in striatum

To investigate the potential role of SAPAP3 in postsynaptic assembly, we examined the levels of a select group of PSD proteins that directly or indirectly interact with SAPAPs, using biochemically purified PSD fractions from the striatum of wild-type and *Sapap3*^{-/-} mice. We found no significant differences in the levels of PSD95, PSD93 or Shank proteins in the striatal PSD between the genotypes (Fig. 4a, b). However, the level of NR1, the obligatory subunit of the NMDAR, was significantly increased (Fig. 4a, b). This result is consistent with our data from extracellular recordings in the striatum, which showed increased NMDAR field potential responses. Although we cannot exclude the possibility that the larger NMDAR-dependent fEPSPs include a contribution from activated conductances other than NMDARs (a limitation of extracellular recordings), our electrophysiological data in combination with these biochemical findings indicate that an increase in NMDARs in the PSD may result in increased synaptic NMDAR activity in *Sapap3*^{-/-} mice. Interestingly, we also found that the level of NR2B, the 'juvenile' subunit of the NMDAR, was significantly increased in the PSD of *Sapap3*^{-/-} mice, whereas NR2A, the 'adult' subunit, was significantly reduced (Fig. 4a, b). In contrast, we found no significant changes in the overall levels of NR1, NR2A and NR2B subunits in the total striatal lysate (Supplementary Fig. 3a, b), suggesting that the changes in these subunits occur exclusively at the PSD. Synaptic NMDARs are known to

undergo a developmental switch from predominantly NR1/NR2B to NR1/NR2A or NR1/NR2A/NR2B during the first two postnatal weeks of life^{22–27}. Although the underlying molecular mechanisms of this developmental switch are not well understood, postsynaptic scaffolding proteins, including the PSD95 and SAPAP family of proteins, have been implicated in this process^{28–31}. Our data indicate that the NMDAR subunit composition at cortico-striatal synapses of adult *Sapap3*^{-/-} mice is similar to that of wild-type juvenile mice^{27,28}, indicating that SAPAP3 may play an important part in synaptic NMDAR subunit switching and maturation at cortico-striatal synapses.

Structure of cortico-striatal synapses in *Sapap3*^{-/-} mice

To determine whether these electrophysiological and biochemical defects are accompanied by morphological changes, we examined the spine density of medium spiny neurons in the striatum using both Golgi staining and DiI filling by Diolistics. We found similar spine densities on medium spiny neurons in *Sapap3*^{-/-} and wild-type mice at postnatal day (P)21 (Supplementary Fig. 3c–e) and in adults (data not shown), indicating that spine formation and maintenance were not affected by the lack of SAPAP3 proteins.

We next examined whether deletion of *Sapap3* affects PSD ultrastructure at cortico-striatal synapses using electron microscopic analysis to measure the length and thickness of the PSD in the striatum (Fig. 4c, d). We found no significant difference in length of the PSD between *Sapap3*^{-/-} and wild-type mice (Fig. 4e). The PSD is reported to exhibit a laminar organization: NMDARs, scaffolding proteins and signalling proteins reside in the electron-dense layer near the synaptic membrane, whereas proteins linked to trafficking and the actin cytoskeleton reside in the filamentous fringe towards the cytoplasm³². Accordingly, we independently measured the thickness of the 'dense layer' and the fringe 'light layer'. We found a small but significant reduction in thickness of the dense layer but not the light layer (Fig. 4f, g) of the PSD in striatal synapses in *Sapap3*^{-/-} mice, suggesting a subtle defect in the structure of the postsynaptic complex. This finding is consistent with the proposed role of SAPAP3 as a postsynaptic scaffolding protein at excitatory synapses^{12,17}.

Rescue by striatal-specific expression of *Sapap3*

Our findings of altered cortico-striatal synaptic function and structure in *Sapap3*-mutant mice, along with pre-existing evidence for the role of striatal circuitry in the pathogenesis of OC-spectrum disorders led us to ask whether loss of SAPAP3 in the striatum was critical for the phenotypes of *Sapap3*^{-/-} mice. We therefore investigated whether selective expression of SAPAP3 in the striatum was sufficient to rescue the phenotypes of *Sapap3*^{-/-} mice. We generated lentiviral vectors that express either green fluorescent protein (GFP) alone or a GFP–SAPAP3 fusion protein. GFP was fused to SAPAP3 at the amino terminus, a region with no known protein–protein interaction sites or motifs. As expected, GFP–SAPAP3 fusion proteins properly localized to synapses when transfected into cultured neurons (Supplementary Fig. 4a).

Lentiviruses expressing either GFP or GFP–SAPAP3 were delivered to the striatum of *Sapap3*^{-/-} mice by bilateral microinjections at P7. To minimize injury and maximize lentiviral infection, we gave injections at one anterior and one posterior site per hemisphere and delivered lentivirus to 8 locations per hemisphere by injection of lentivirus along the needle path (Fig. 5a; see Methods). Sustained expression of GFP and GFP–SAPAP3 was observed in both dorsal and ventral striatum in post-mortem examination of brains at the end of the study (Fig. 5b–e, and Supplementary Fig. 4b, c).

At 4–6 months after injection, behavioural analyses were conducted to assess grooming, facial lesions and anxiety-like behaviours. When examined by continuous videotaping, excessive grooming was markedly reduced in *Sapap3*^{-/-} mice injected with lentivirus encoding GFP–SAPAP3 in comparison to *Sapap3*^{-/-} mice injected with the control GFP lentivirus (Fig. 5f). Moreover, facial

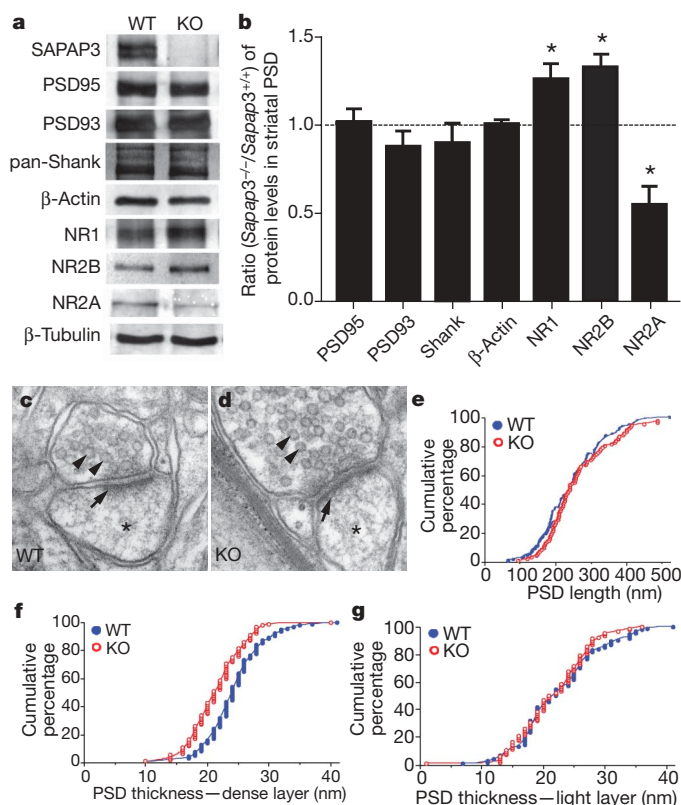


Figure 4 | Structural and biochemical analyses of cortico-striatal synapses in *Sapap3*-mutant mice. **a, b**, The levels of PSD95 (DLG4), PSD93 (DLG2) and Shank (SHANK1-3) in striatal PSD fractions are not affected in *Sapap3*^{-/-} mice. The levels of NR1 (GRIN1) and NR2B (GRIN2B) subunits are increased, whereas that of NR2A (GRIN2A) is decreased. β-actin and β-tubulin were used as loading controls. **P* < 0.05, two-tailed *t*-test. All data are presented as means ± s.e.m. **c, d**, Electron micrographs show the presence of synaptic vesicles (arrowheads), postsynaptic densities (arrows) and dendritic spines (asterisks). **e**, The length of the PSD is not significantly different in wild-type and *Sapap3*^{-/-} mice. **f, g**, The thickness of the dense layer (**f**) of the PSD in *Sapap3*^{-/-} mice, but not the light layer (**g**), is reduced. *P* < 0.001, two-tailed *t*-test; *n* = 94 for wild type and *n* = 92 for *Sapap3*^{-/-} mice.

lesion severity was also significantly reduced in *Sapap3*^{-/-} mice injected with GFP-*Sapap3* lentivirus (Fig. 5g, h). All eight *Sapap3*^{-/-} mice injected with the control GFP lentivirus developed open skin lesions (six with large lesions, one with medium, and one with small open wounds). In contrast, only three of the eight *Sapap3*^{-/-} mice injected with GFP-SAPAP3 developed open skin lesions (one with large, one with medium, and one with small open wounds), whereas two had small hairless patches and three had normal skin conditions (Fig. 5i). Furthermore, *Sapap3*^{-/-} mice injected with lentivirus encoding GFP-SAPAP3 also showed reduced anxiety-like behaviour in the dark-light emergence test (Fig. 5j, k), with no change in total activity (data not shown). Finally, lentiviral-mediated expression of GFP-SAPAP3 in the striatum also rescued defects in cortico-striatal synaptic transmission in *Sapap3*^{-/-} mice (Fig. 5l, m). Together, these data show that early postnatal expression of GFP-SAPAP3 selectively in the striatum is sufficient to alleviate the major manifestations of the OCD-like phenotype in *Sapap3*^{-/-} mice.

Discussion

Although dysfunction of cortico-striato-thalamo-cortical circuitry has been widely implicated in OC-spectrum disorders, the nature of this dysfunction remains unclear. Our study of *Sapap3*-mutant mice has identified an important role for SAPAP3 at cortico-striatal synapses and, although there are inherent limitations of evaluating thought content in mice, our study further indicates that cortico-striatal synaptic defects may be central to the genesis of OCD-like behaviours. These conclusions are supported by several lines of evidence. First, *Sapap3*^{-/-} mice exhibit OCD-like behaviours: excessive time spent performing a ritualistic action to the point of being self-injurious, increased anxiety-like behaviours, and response of

these manifestations to a selective serotonin reuptake inhibitor. Second, SAPAP3 protein is localized to excitatory, but not inhibitory, synapses¹⁵. Third, *Sapap3* is the only member of the *Sapap* family that is highly expressed in the striatum. Fourth, *Sapap3*-mutant mice have reduced cortico-striatal synaptic transmission and an NMDAR subunit composition suggestive of immature cortico-striatal synapses. Finally, selective expression of *Sapap3* in the striatum of *Sapap3*^{-/-} mice rescues the synaptic defects and OCD-like behaviours.

Anxiety is attributed primarily to dysfunction of the amygdala and ventral hippocampus. Interestingly, SAPAP3, unlike SAPAP1 and SAPAP4, is not highly expressed in the amygdala (Supplementary Fig. 5a–d), whereas all four members of the SAPAP family are expressed in the ventral hippocampus (Supplementary Fig. 5e–h), suggesting that anxiety-like behaviour in *Sapap3*^{-/-} mice is unlikely to be due to the lack of SAPAP3 in amygdala or ventral hippocampus. The unique expression of SAPAP3 in the striatum and the alleviation of the anxiety-like behaviour by lentiviral-mediated expression of GFP-SAPAP3 in the striatum suggest that the anxiety-like behaviour in *Sapap3*^{-/-} mice originates from striatal defects. Because the amygdala projects to the striatum, it is also possible that some functions of amygdalar-striatal projections are impaired in *Sapap3*^{-/-} mice. Additionally, our video surveillance data indicate that *Sapap3*-mutant mice have a disrupted sleep pattern (Supplementary Fig. 6), which is alleviated by fluoxetine treatment (Supplementary Fig. 7). This sleep disturbance may also contribute to the anxiety-like behaviour in *Sapap3*^{-/-} mice.

A principal role of the striatum is to integrate the various inputs arriving from the cortex and use this information to select certain motor and/or cognitive programs (decision making), which are subsequently carried out through the direct and indirect pathways of

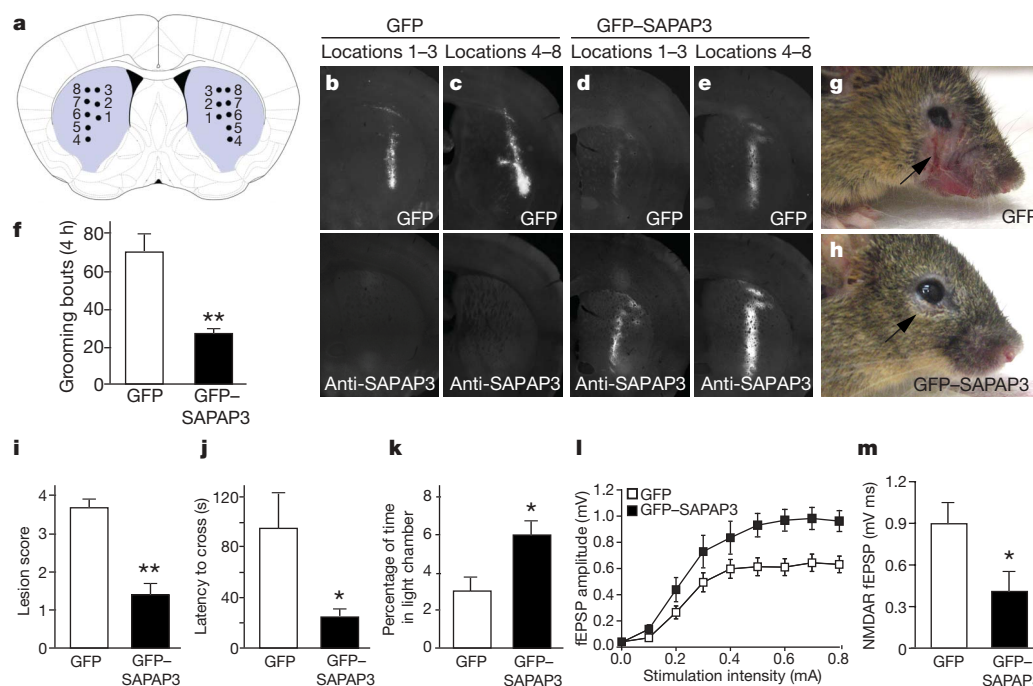


Figure 5 | Lentiviral-mediated rescue of behavioural and synaptic defects in *Sapap3*-mutant mice. **a**, Diagram showing the approximate locations of microinjections in the striatum of *Sapap3*^{-/-} mice. Injection site 1, locations 1–3 are more anterior than injection site 2, locations 4–8. **b–c**, Brain sections from a *Sapap3*^{-/-} mouse injected with GFP lentivirus show GFP fluorescence (**b**, **c**, upper panels) and an absence of SAPAP3 staining (**b**, **c**, lower panels). **d–e**, Brain sections from a *Sapap3*^{-/-} mouse injected with lentivirus encoding GFP-SAPAP3 show both GFP fluorescence (**d**, **e**, upper panels) and SAPAP3 immunostaining (**d**, **e**, lower panels). **f**, Compared with *Sapap3*^{-/-} mice with GFP, *Sapap3*^{-/-} mice with GFP-SAPAP3 showed significantly reduced over-grooming behaviour. ***P* < 0.01, two-tailed *t*-test; *n* = 8 mice per group for **f**, **i–k**. **g**, **h**, *Sapap3*^{-/-} mice with GFP-SAPAP3 (**h**) had reduced

severity of facial lesions when compared to *Sapap3*^{-/-} mice with GFP (**g**). **i**, Semi-quantitative lesion scores. ***P* < 0.01, Mann-Whitney U test. **j**, **k**, Reduced anxiety-like behaviours in *Sapap3*^{-/-} mice injected with GFP-*Sapap3* lentivirus in the dark-light emergence test, including decreased latency to cross from the dark to light chamber (**j**) and increased time in the light chamber (**k**). **P* < 0.05, two-tailed *t*-test. **l**, **m**, Field recordings from infected striatal areas of P21–P25 *Sapap3*^{-/-} mice with GFP-SAPAP3 showed an increase in cortico-striatal fEPSP amplitude (**l**) and a reduction of NMDAR-dependent fEPSP area (**m**). *P* < 0.001, repeated measures ANOVA for **l**; **P* < 0.05, two-tailed *t*-test for **m**; *n* is 12 and 10 for **l**, and 10 and 9 for **m** for GFP-injected and GFP-SAPAP3 injected mice, respectively. All data are presented as means ± s.e.m.

the basal ganglia⁴. A prominent model for OCD is that an activity imbalance between the direct and indirect pathways leads to behavioural abnormalities of OCD³, and recent studies have shown that the two pathways are differentially regulated at cortico-striatal synapses^{33–35}. We have identified defects in both AMPAR- and NMDAR-dependent synaptic transmission at cortico-striatal synapses. These data are from extracellular field recordings, representing a population average of the effects on individual neurons. If the observed effects are differentially manifested by medium spiny neurons in the direct or indirect pathway, an imbalance in activity could be readily envisioned.

Studies of the mechanisms of OC-spectrum disorders have mostly focused on the serotonergic and dopaminergic systems. However, recent genetic association studies of OCD in humans have implicated genes important for glutamatergic neurotransmission^{36–38}. Together with our study of SAPAP3 in mice, these observations raise the possibility that defects in excitatory synaptic transmission in the cortico-striatal circuit may contribute to the pathogenesis of OC-spectrum disorders in humans.

METHODS SUMMARY

Behavioural analysis. *Sapap3*-knockout mice were generated by homologous recombination in R1 embryonic stem cells using standard procedures³⁹. Adult mice, age 4–8 months, were used for behavioural analyses. For analyses of grooming behaviours⁴⁰, habituated, individually housed animals were videotaped for 24 h under 700 lx (day) and ~2 lx (red light at night) illumination. Tests for anxiety-like behaviours were performed as described^{41,42}. All experiments were done blind to genotypes. Because of the presence of obvious facial lesions in *Sapap3*-mutant mice, it is impossible to perform video decoding in a blind manner. To avoid bias, trained individuals unfamiliar with the project (listed in the Acknowledgements) were recruited to decode the videos.

Cortico-striatal electrophysiology. Acute sagittal brain slices (300 µm) from P17–P25 mice were used for all experiments. The field recording electrode was placed in the dorso-lateral striatum and a monopolar stimulation electrode was placed in the corpus callosum. All recordings were performed at 30–32 °C and in the presence of picrotoxin and all data were collected and analysed before unblinding of genotypes.

Stereotaxic injection. One-week-old *Sapap3*^{−/−} mice were anaesthetized and placed in a mouse head holder. GFP lentivirus or GFP-*Sapap3* lentivirus was bilaterally injected into the striatum through 2 sites at 8 locations per hemisphere. Behavioural analyses were performed 4–6 months after injection.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 May; accepted 16 July 2007.

- Karno, M., Golding, J. M., Sorenson, S. B. & Burnam, M. A. The epidemiology of obsessive-compulsive disorder in five US communities. *Arch. Gen. Psychiatry* **45**, 1094–1099 (1988).
- Torres, A. R. et al. Obsessive-compulsive disorder: prevalence, comorbidity, impact, and help-seeking in the British national psychiatric morbidity survey of 2000. *Am. J. Psychiatry* **163**, 1978–1985 (2006).
- Swedo, S. E. & Snider, L. A. In *Neurobiology of Mental Illness* (eds Nestler, E. J. & Charney, D. S.) 628–638 (Oxford Univ. Press, New York, 2004).
- Graybiel, A. M. & Rauch, S. L. Toward a neurobiology of obsessive-compulsive disorder. *Neuron* **28**, 343–347 (2000).
- Aouizerate, B. et al. Pathophysiology of obsessive-compulsive disorder: a necessary link between phenomenology, neuropsychology, imagery and physiology. *Prog. Neurobiol.* **72**, 195–221 (2004).
- Hanna, G. L. et al. Genome-wide linkage analysis of families with obsessive-compulsive disorder ascertained through pediatric probands. *Am. J. Med. Genet.* **114**, 541–552 (2002).
- Shugart, Y. Y. et al. Genomewide linkage scan for obsessive-compulsive disorder: evidence for susceptibility loci on chromosomes 3q, 7p, 1q, 15q, and 6q. *Mol. Psychiatry* **11**, 763–770 (2006).
- Nestadt, G. et al. A family study of obsessive-compulsive disorder. *Arch. Gen. Psychiatry* **57**, 358–363 (2000).
- Inouye, E. Similar and dissimilar manifestations of obsessive-compulsive neurosis in monozygotic twins. *Am. J. Psychiatry* **121**, 1171–1175 (1965).
- Carey, G. & Gottesman, I. I. In *Anxiety: New Research and Changing Concepts* (eds Klein, D. F. & Rabkin, J.) 117–136 (Raven Press, New York, 1981).
- Chakrabarty, K., Bhattacharyya, S., Christopher, R. & Khanna, S. Glutamatergic dysfunction in OCD. *Neuropsychopharmacology* **30**, 1735–1740 (2005).
- Kim, E. et al. GKAP, a novel synaptic protein that interacts with the guanylate kinase-like domain of the PSD-95/SAP90 family of channel clustering molecules. *J. Cell Biol.* **136**, 669–678 (1997).
- Takeuchi, M. et al. SAPAPs. A family of PSD-95/SAP90-associated proteins localized at postsynaptic density. *J. Biol. Chem.* **272**, 11943–11951 (1997).
- Scannevin, R. H. & Huganir, R. L. Postsynaptic organization and regulation of excitatory synapses. *Nature Rev. Neurosci.* **1**, 133–141 (2000).
- Kim, E. & Sheng, M. PDZ domain proteins of synapses. *Nature Rev. Neurosci.* **5**, 771–781 (2004).
- Funke, L., Dakoji, S. & Brecht, D. S. Membrane-associated guanylate kinases regulate adhesion and plasticity at cell junctions. *Annu. Rev. Biochem.* **74**, 219–245 (2005).
- Welch, J. W., Wang, D. & Feng, G. Differential mRNA expression and protein localization of the SAP90/PSD-95-associated proteins (SAPAPs) in the nervous system of the mouse. *J. Comp. Neurol.* **472**, 24–39 (2004).
- Kindler, S., Rehbein, M., Classen, B., Richter, D. & Bockers, T. M. Distinct spatiotemporal expression of SAPAP transcripts in the developing rat brain: a novel dendritically localized mRNA. *Brain Res. Mol. Brain Res.* **126**, 14–21 (2004).
- Malinow, R. & Malenka, R. C. AMPA receptor trafficking and synaptic plasticity. *Annu. Rev. Neurosci.* **25**, 103–126 (2002).
- Prybylowski, K. & Wenthold, R. J. N-Methyl-D-aspartate receptors: subunit assembly and trafficking to the synapse. *J. Biol. Chem.* **279**, 9673–9676 (2004).
- Nicoll, R. A., Tomita, S. & Brecht, D. S. Auxiliary subunits assist AMPA-type glutamate receptors. *Science* **311**, 1253–1256 (2006).
- Sheng, M., Cummings, J., Roldan, L. A., Jan, Y. N. & Jan, L. Y. Changing subunit composition of heteromeric NMDA receptors during development of rat cortex. *Nature* **368**, 144–147 (1994).
- Shi, J., Aamodt, S. M. & Constantine-Paton, M. Temporal correlations between functional and molecular changes in NMDA receptors and GABA neurotransmission in the superior colliculus. *J. Neurosci.* **17**, 6264–6276 (1997).
- Stocca, G. & Vicini, S. Increased contribution of NR2A subunit to synaptic NMDA receptors in developing rat cortical neurons. *J. Physiol.* **507**, 13–24 (1998).
- Tovar, K. R. & Westbrook, G. L. The incorporation of NMDA receptors with a distinct subunit composition at nascent hippocampal synapses *in vitro*. *J. Neurosci.* **19**, 4180–4188 (1999).
- Chapman, D. E., Keefe, K. A. & Wilcox, K. S. Evidence for functionally distinct synaptic NMDA receptors in ventromedial versus dorsolateral striatum. *J. Neurophysiol.* **89**, 69–80 (2003).
- Li, L., Murphy, T. H., Hayden, M. R. & Raymond, L. A. Enhanced striatal NR2B-containing methyl-D-aspartate receptor-mediated synaptic currents in a mouse model of Huntington disease. *J. Neurophysiol.* **92**, 2738–2746 (2004).
- Sans, N. et al. A developmental change in NMDA receptor-associated proteins at hippocampal synapses. *J. Neurosci.* **20**, 1260–1271 (2000).
- Barria, A. & Malinow, R. Subunit-specific NMDA receptor trafficking to synapses. *Neuron* **35**, 345–353 (2002).
- Prybylowski, K. et al. The synaptic localization of NR2B-containing NMDA receptors is controlled by interactions with PDZ proteins and AP-2. *Neuron* **47**, 845–857 (2005).
- van Zundert, B., Yoshii, A. & Constantine-Paton, M. Receptor compartmentalization and trafficking at glutamate synapses: a developmental proposal. *Trends Neurosci.* **27**, 428–437 (2004).
- Valtschanoff, J. G. & Weinberg, R. J. Laminar organization of the NMDA receptor complex within the postsynaptic density. *J. Neurosci.* **21**, 1211–1217 (2001).
- Day, M. et al. Selective elimination of glutamatergic synapses on striatopallidal neurons in Parkinson disease models. *Nature Neurosci.* **9**, 251–259 (2006).
- Kreitzer, A. C. & Malenka, R. C. Endocannabinoid-mediated rescue of striatal LTD and motor deficits in Parkinson's disease models. *Nature* **445**, 643–647 (2007).
- Surmeier, D. J., Ding, J., Day, M., Wang, Z. & Shen, W. D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. *Trends Neurosci.* **30**, 228–235 (2007).
- Arnold, P. D. et al. Association of a glutamate (NMDA) subunit receptor gene (*GRIN2B*) with obsessive-compulsive disorder: a preliminary study. *Psychopharmacology* **174**, 530–538 (2004).
- Arnold, P. D., Sicard, T., Burroughs, E., Richter, M. A. & Kennedy, J. L. Glutamate transporter gene *SLC1A1* associated with obsessive-compulsive disorder. *Arch. Gen. Psychiatry* **63**, 769–776 (2006).
- Dickel, D. E. et al. Association testing of the positional and functional candidate gene *SLC1A1/EAAC1* in early-onset obsessive-compulsive disorder. *Arch. Gen. Psychiatry* **63**, 778–785 (2006).
- Feng, G. et al. Dual requirement for gephyrin in glycine receptor clustering and molybdoenzyme activity. *Science* **282**, 1321–1324 (1998).
- Greer, J. M. & Capecchi, M. R. Hoxb8 is required for normal grooming behavior in mice. *Neuron* **33**, 23–34 (2002).
- Pogorelov, V. M., Rodriguez, R. M., Inscio, M. L., Caron, M. G. & Wetsel, W. C. Novelty seeking and stereotypic activation of behavior in mice with disruption of the *DAT1* gene. *Neuropsychopharmacology* **30**, 1818–1831 (2005).
- Weisstaub, N. V. et al. Cortical 5-HT_{2A} receptor signaling modulates anxiety-like behaviors in mice. *Science* **313**, 536–540 (2006).
- Bakeman, R. & Gottman, J. M. In *Observing Interaction: An Introduction to Sequential Analyses* 56–90 (Cambridge Univ. Press, New York, 1997).
- Treit, D. & Fundytus, M. Thigmotaxis as a test for anxiolytic activity in rats. *Pharmacol. Biochem. Behav.* **31**, 959–962 (1988).

45. Crawley, J. N. & Goodwin, F. K. Preliminary report of a simple animal behavior model for the anxiolytic effects of benzodiazepines. *Pharmacol. Biochem. Behav.* **12**, 167–170 (1980).
46. Feng, G. *et al.* Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron* **28**, 41–51 (2000).
47. Gan, W. B., Grutzendler, J., Wong, W. T., Wong, R. O. & Lichtman, J. W. Multicolor “DiOlistic” labeling of the nervous system using lipophilic dye combinations. *Neuron* **27**, 219–225 (2000).
48. Lois, C., Hong, E. J., Pease, S., Brown, E. J. & Baltimore, D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295**, 868–872 (2002).
49. Parker, M. J., Zhao, S., Brecht, D. S., Sanes, J. R. & Feng, G. PSD93 regulates synaptic stability at neuronal cholinergic synapses. *J. Neurosci.* **24**, 378–388 (2004).
50. Lau, L. F. & Huganir, R. L. Differential tyrosine phosphorylation of *N*-methyl-D-aspartate receptor subunits. *J. Biol. Chem.* **270**, 20036–20041 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Gross, K. Phend and L. Qiu for technical assistance, and L. Phillips, L. Nguyen, S. Greeter, J. Wilkins and M. Fukui for

assistance in behavioural testing and decoding of video tapes. We thank M. Ehlers for the anti-NR2B antibody and E. Kim for the anti-Shank antibody. We also thank M. Caron, M. Ehlers, Z. He, J. Sanes, F. Wang, A. West and members of the Feng laboratory for critical reading of the manuscript. This work was supported by grants from NINDS and NIMH to G.F., R.J.W. and N.C.; by unrestricted funds to W.C.W.; and by the Intramural Research Program of NIEHS to S.M.D. J.M.W. was supported by an NSF pre-doctoral fellowship and an NIH National Research Service Award. N.C. is a recipient of a Klingenstein Fellowship in the Neurosciences and a NARSAD Young Investigator Award. G.F. is a recipient of a Sloan Fellowship, a Klingenstein Fellowship in the Neurosciences, an EJLB Foundation Scholar Research Program Award, a McKnight Neuroscience of Brain Disorders Award and a Hartwell Foundation Individual Biomedical Research Award.

Author Contributions J.M.W., J. Lu, R.M.R., N.C.T., J.P., J.-D.D., C.F., M.C. and J.P.A. participated in the design, analysis and execution of experiments. G.F., N.C., W.C.W., J.M.W., R.J.W., S.M.D. and J. Luo participated in the design, analysis and interpretation of experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.F. (feng@neuro.duke.edu).

METHODS

Mice. A targeting vector was designed to replace exon 3 (containing the translation initiation codon) of the *Sapap3* gene with a *NEO* cassette. Genotypes were determined by PCR of mouse tail DNA, using primer F1 (ATTGGTAGG-CAATACCAACAGG) and R1 (GCAAAGGCTCTTCATATTGTTGG) for the wild-type allele (147 base pairs), and F1 and R2 (CTTTGTGGTCTAA-GTACTGTGG; in neo cassette) for the mutant allele (222 base pairs). Primer locations are indicated in Supplementary Fig. 1.

Grooming behaviour⁴⁰. Because behaviours were most representative at the times 02:00–06:00, 10:00–14:00 and 18:00–22:00, these segments were selected for analyses using Noldus Observer software (Leesburg). The total amount of time in each 4-h segment spent grooming, eating, sleeping, rearing, digging or pushing bedding (shifting) was determined, along with general states of activity. Grooming included all sequences of face-wiping, scratching/rubbing of head and ears, and full-body grooming. Bouts of behaviour lasted at least 3 s; pauses longer than 3 s constituted a new bout. Observer reliability was determined with Cohen's kappa⁴³.

Anxiety-like behaviours^{41,42}. Zero maze: an elevated zero maze⁴¹ was indirectly illuminated at 60 lx. Testing commenced with an animal being introduced into a closed area of the maze. Behaviour was video-taped for 5 min and subsequently scored by trained observers using the Noldus Observer (Nodus Information Technology). Anxiety-like behaviour was deduced on the basis of the percentage of time spent in the open areas and the latency to enter the open areas. Total activity time in the maze was an indicator of motor activity. Observer reliability was determined with Cohen's Kappa.

Open field: spontaneous locomotor activity was evaluated over 30 min in an automated Omnitech Digiscan apparatus (AccuScan Instruments) as described⁴⁴. Locomotor activity was assessed as total distance travelled (cm). Anxiety-like behaviour was defined by the percentage of time spent in the centre as contrasted with the percentage of time spent in the perimeter (thigmotaxis) of the open field.

Dark–light emergence test⁴⁵. Mice were adapted in an adjacent room to low light conditions (~40 lx) and the test room was initially under similar illumination. Testing was conducted in a two-chambered test apparatus (Med-Associates), with one side draped in black cloth (for example, dark-chamber) and the other illuminated at ~1,400 lx (for example, light-chamber) with a 170 mA high intensity house light and overhead fluorescent lamps. Immediately on placing the mice into the darkened chamber, the lighted chamber was illuminated and the door between the two chambers was opened. The mice were allowed to freely explore the apparatus for 5 min. The latency to emerge from the darkened into the lighted chamber and the percentage of time spent in the illuminated chamber were used as indices of anxiety-like behaviours. Total motor activity in the two chambers was measured by infrared beam-breaks.

Fluoxetine treatment. Adult mice 4–6 months of age were injected with fluoxetine (5 mg kg⁻¹, intraperitoneally) once a day for one or six days and behaviours were tested 24 h after the final injection. Grooming and anxiety-like behaviours were tested as described above.

Skin analyses. Haematoxylin and eosin staining was performed on facial skin below the eyes by following a standard protocol. For visualization of sensory nerves in skin, *Sapap3*-mutant mice were crossed with Thy1GFP-J mice⁴⁶. Mice were perfused transcardially with saline and 4% paraformaldehyde, and skin from just below the eyes was dissected and post-fixed for 24 h. Confocal images were taken from 50-μm cryosections. Skin lesions were scored as follows: 0, normal skin; 1, hairless patches; 2, small open-wound lesions, length <2 mm; 3, moderate open-wound lesions, length 2–4 mm; 4, large open-wound lesions, length >4 mm.

Cortico-striatal electrophysiology. Recording perfusion solution contained (in mM): 119 NaCl, 2.5 KCl, 1.2 NaH₂PO₄, 26 NaHCO₃, 1 MgCl₂, 2 CaCl₂ and 0.1 picrotoxin (GABA_A receptor antagonist). Slicing and recovery solutions were identical to perfusion solution except for containing 0.5 CaCl₂ and no picrotoxin. APV (50 μM) and NBQX (50 μM) were obtained from Tocris. Solutions were continuously equilibrated with 95% O₂ and 5% CO₂ (pH 7.4) and perfusion flow rate was 2 ml min⁻¹. Slices were allowed to recover for a minimum of 1 h at 30–32 °C, following slicing. All recordings were performed at 30–32 °C.

The field recording electrode was placed in the dorso-lateral striatum and a monopolar stimulation electrode was placed in the corpus callosum. Current was delivered to the stimulating electrode using an AMPI. Stimulus Isolator (AMPI) for 150 μs. Three distinct components were resolved in the majority of recordings: stimulation artefact, negative peak 1 (NP1, action-potential-derived on the basis of latency, resistance to NBQX and picrotoxin, and sensitivity to tetrodotoxin) and negative peak 2 (NP2, fEPSP based on latency and sensitivity to NBQX; in addition, sensitivity to tetrodotoxin indicates the response is not due to direct activation by stimulating electrode current). The callosal stimulation site was chosen over an intra-striatal site to minimize

activation of non-cortical axons. Data were acquired at 20 kHz and filtered at 2 kHz using MultiClamp 700B amplifier and pClamp 10.0 software (Axon Instruments). Data were analysed offline using Clampfit 10.0 (Axon Instruments). Five consecutive responses were averaged before measuring amplitude, slope or area in the respective assays. When amplitudes are reported, similar conclusions were obtained by slope analysis. Paired-pulse responses were evoked using a stimulation intensity that yielded the maximal fEPSP response. Slope values used for paired-pulse ratios refer to slope during the period from 20–80% of the peak response. NMDAR fEPSPs were evoked using the same stimulation intensity that yielded the maximal fEPSP response under the basal recording conditions used to generate the input–output curves. The area of NMDAR field potential responses was measured during a standard 20-ms time window beginning approximately 8 ms after stimulation. All data were collected and analysed before unblinding of genotypes. In addition, NMDAR field potential recordings and correlative PSD biochemical studies were performed in two different laboratories and were unblinded at the same time.

For *in vivo* viral expression rescue experiments, the same slicing, recording, and analysis methods were used as described above except that the stimulating and recording electrodes were placed within the fluorescent green tissue. Only slices where green cells were visible in the dorsolateral recording region were used. Viral rescue experiments were performed and analysed blind to the viral identity.

Diolistics. 'Bullets' were prepared by coating tungsten particles (Bio-rad) with DiI (Molecular Probes-Invitrogen) dissolved in methylene chloride as described⁴⁷. Mice were perfused transcardially, first with normal saline, and then with buffered 4% paraformaldehyde. Fixed mice were decapitated and the head was allowed to post-fix for 24–48 h in buffered 4% paraformaldehyde at 4 °C. Brains were dissected and 250 μm coronal slices were cut on a vibratome.

Diolistic filling of individual cells was performed using a standard Bio-Rad diolistic system. Slices were placed into the interior chamber of a Falcon organ tissue culture dish (catalogue number 3,037) and excess buffer was removed. A Millipore 3.0-μm Isopore filter (catalogue number TSTP04704) was placed to rest on top of the tissue culture dish and a Kimwipe was placed on top of the Millipore filter. Bullets were shot through the Kimwipe and filter barrier using a helium charge of 552 kPa (80 psi). Individual labelling of 3–15 striatal medium spiny neurons was successful in approximately 50% of slices. Following the diolistic procedure, slices were allowed to incubate in buffer at 4 °C for 1–2 h. Following incubation, slices were immediately mounted and imaged on a confocal microscope.

Lentivirus production and stereotaxic injection. The lentivirus vector was based on FUGW⁴⁸ and modified by replacing the ubiquitin promoter with a CMV promoter to improve neuronal expression. GFP was fused to the amino terminus of SAPAP3 to generate the GFP–SAPAP3 fusion protein. The complementary DNA clone encoding GFP or GFP–SAPAP3 was placed downstream of the CMV promoter between *Bam*HI and *Eco*RI sites⁴⁸. Viral particles were produced by transient transfection of 293T cells with the lentiviral vector encoding GFP or GFP–SAPAP3 along with the envelope (VSVg) and packaging (Δ8.9) vectors. High titre viral suspensions were obtained by ultra-centrifugation at 77,000g for 90 min.

One-week-old *Sapap3*^{-/-} mice were anaesthetized with a mixture of ketamine and xylazine (intraperitoneally), and placed in a mouse head holder (David Kopf Instruments). Lentivirus encoding GFP or GFP–SAPAP3 was bilaterally injected into the striatum through two sites at eight locations per hemisphere. At each of the injection sites, the microinjection needle was advanced to the deepest (ventral) position for the first injection; additional injections were made every 0.3 mm while withdrawing the injection needle. The coordinates from Bregma were: injection site 1, location 1–3: anterior 0.5 mm, mediolateral 1.4 mm, dorsoventral 2.8, 2.5, 2.2 mm; injection site 2, location 4–8: anterior 0.2 mm, mediolateral 1.7 mm, dorsoventral 3.3, 3.0, 2.7, 2.4, 2.1 mm. For each injection location, 90 nl of virus was injected using Nanoject II (Drummond Scientific), and the needle was left in place for 3 min after each injection.

Electron microscopic analysis. Striatal tissues for electron microscopic analysis were processed and quantitatively analysed according to protocols previously described³². Synapses with clear membranes (likely to be cut orthogonal to the plane of the synaptic membrane) were photographed at ×15,000 magnification. Data collection and analysis were performed by observers blinded to genotypes. For every clearly defined synapse, data for PSD length and PSD thickness were collected.

PSD preparation and western blot. PSD fractions of the striatum were prepared as described^{17,49}, separated by SDS–PAGE and probed with specific antibodies. The relative amount of β-tubulin and β-actin were used as loading controls for quantification. Antibodies for SAPAP3, PSD95, PSD93, Shank and NR2B have been described previously^{17,49,50}. Antibodies for NR1 (mouse monoclonal) and NR2A (rabbit polyclonal) were from BD Biosciences and Upstate, respectively.

Gap junction adhesion is necessary for radial migration in the neocortex

Laura A. B. Elias^{1,2}, Doris D. Wang^{1,2} & Arnold R. Kriegstein²

Radial glia, the neuronal stem cells of the embryonic cerebral cortex, reside deep within the developing brain and extend radial fibres to the pial surface, along which embryonic neurons migrate to reach the cortical plate. Here we show that the gap junction subunits connexin 26 (Cx26) and connexin 43 (Cx43) are expressed at the contact points between radial fibres and migrating neurons, and acute downregulation of Cx26 or Cx43 impairs the migration of neurons to the cortical plate. Unexpectedly, gap junctions do not mediate neuronal migration by acting in the classical manner to provide an aqueous channel for cell–cell communication. Instead, gap junctions provide dynamic adhesive contacts that interact with the internal cytoskeleton to enable leading process stabilization along radial fibres as well as the subsequent translocation of the nucleus. These results indicate that gap junction adhesions are necessary for glial-guided neuronal migration, raising the possibility that the adhesive properties of gap junctions may have an important role in other physiological processes and diseases associated with gap junction function.

During neocortical brain development, radial glia, which reside in the ventricular zone and extend a fibre to the pial surface, serve two purposes: (1) as stem cells of the developing cortex, giving rise to neurons^{1,2}, and (2) as guides along which the neurons migrate to reach the correct lamina of the cortical plate, where they will become pyramidal cells in the adult cortex^{3–5}. Interstitial junctions containing filamentous material have been found between migrating neurons and radial fibres, suggesting a close association between the two cell types^{6–8}.

The molecules that mediate the critical interaction between the radial glial fibre and the migrating neuron remain largely unknown. Astrotactin has been described as an adhesion molecule in the cerebellar cortex^{9,10}, but has not been implicated in neocortical development. Neuregulin and its soluble form, glial growth factor, mediate bi-directional signalling between the radial fibre and the migrating neuron, and are important for radial glial maintenance and neuronal migration¹¹. $\alpha_3\beta_1$ integrin expressed in migrating neurons is involved in glial recognition, and α_v integrin expressed in radial glia is involved in neuronal adhesion¹². The expression of gap junctions in radial fibres and migrating neurons, together with the detection of gap junctions between nestin⁺ radial glial fibres and nestin[–] cells by electron microscopy, suggests the possibility that gap junctions play a part in neuron–glial interactions and neuronal migration¹³. This hypothesis is supported by the change in the distribution of bromodeoxyuridine (BrdU) pulse-labelled cells in the Cx43 genetic knockout (Cx43KO)¹⁴. Furthermore, strong evidence links Cx43 to migration in neural crest cells and gliomas; however, the mechanisms are not fully understood^{15–18}.

Gap junctions are large-diameter channels made up of two hemichannels—each composed of six connexin subunits—on opposing membranes that join through hydrophobic interactions and form an aqueous pore between the cytoplasm of two adjacent cells¹⁹. The gap junction subunits expressed in the developing cortex include Cx26 (also known as Gjb2), a β_1 connexin family member, and Cx43 (also known as Gja1), an α_1 connexin family member²⁰. There is strong evidence that radial glia are coupled electrically and chemically to each other by gap junction channels during embryonic

development²¹. Functionally, it has been suggested that gap junctions have a role in radial glial cell cycle regulation because pharmacological block of gap junctions, or Ca^{2+} waves mediated by gap junction hemichannels, inhibits the cell cycle^{22,23}. However, the possible role of gap junctions in cortical migration remains largely unexplored.

Gap junction expression at contact points

The gap junction subunits Cx26 and Cx43 are expressed in the developing rat neocortex. Whereas Cx26 is evenly distributed from the ventricular zone through the intermediate zone to the cortical plate, Cx43 is highly expressed at the ventricular surface and its levels are reduced in the cortical plate (Fig. 1a, b, e, f). Cx26 and Cx43 puncta do not colocalize with each other (Supplementary Fig. 1), because they belong to separate families that do not make heterotypic junctions or heteromeric hemichannels²⁴. The expression of both connexins outside of the ventricular zone proliferative region indicates that gap junctions may regulate functions additional to radial glial proliferation. To address this possibility, we determined the pattern of Cx26 and Cx43 expression with respect to β -III tubulin⁺ migrating neurons and vimentin⁺ radial glial fibres. Both gap junction subunits are expressed in migrating neurons and along radial fibres. In fact, Cx26 and Cx43 are highly localized in neurons to the regions of contact with radial fibres, consistent with the idea that gap junctions may have a role in mediating communication between migrating neurons and their radial guides (Fig. 1, and Supplementary Movies 1, 2).

Cx26 and Cx43 are necessary for migration

To investigate the role of gap junctions in cortical development, we developed acute loss-of-function manipulations for each connexin using RNA interference. Short hairpin RNA (shRNA) constructs that produced significant knockdown of rat Cx26 or Cx43 by western blot in Cos-7 monkey cells were selected (Cx26-shRNA or Cx43-shRNA, respectively), and it was confirmed that the same hairpin sequences with three point mutations (Ctrl-shRNA) failed to produce significant knockdown (Supplementary Fig. 2a, b). We confirmed that

¹Neuroscience Graduate Program, ²Institute for Regeneration Medicine, University of California San Francisco, 513 Parnassus Avenue, San Francisco, California 94143, USA.

Cx26-shRNA and Cx43-shRNA were able to knockdown endogenous connexin protein in the developing rat brain (Supplementary Fig. 2c, d).

We next examined the functional effects of Cx26 or Cx43 down-regulation in the intact developing neocortex. *In utero* intraventricular injection and electroporation of Cx26-shRNA or Cx43-shRNA at embryonic day (E)16 resulted in a striking cellular redistribution pattern compared with Ctrl-shRNA at E18 and E21 (Fig. 2a). We quantified this effect by dividing the cortex into five equal areas and determining the fraction of the total number of shRNA-expressing

(GFP⁺) cells in each area (Fig. 2b). At E18, there was a significant change in the fraction of cells in all areas except in the ventricular zone; most clearly there was a loss of cells in the lower and upper cortical plate (Fig. 2b). The effect was more dramatic at E21; the percentage of cells in each area, except the ventricular zone/SVZ (sub-ventricular zone), was significantly different. Most strikingly, there was an accumulation of cells in the intermediate zone, with far fewer cells in the cortical plate (Fig. 2b). The intermediate phenotype of the Cx26-shRNA is due to a rostral caudal gradient such that Cx26-shRNA exerts a stronger effect on neuronal migration in caudal regions where Cx26 is highly expressed (Supplementary Fig. 3). The target-specificity of these results was confirmed by using alternative shRNA targets for Cx26 and Cx43 (Supplementary Fig. 4).

Because radial glia express connexins, it was possible that the defect in neuronal migration was due to a disruption in the radial glial scaffolding. However, immunohistochemical analysis of the radial glia, as well as the migration of wild-type neurons in electroporated regions, seemed normal (Supplementary Fig. 5). To address this possibility further, transplant assays were performed. Brains were

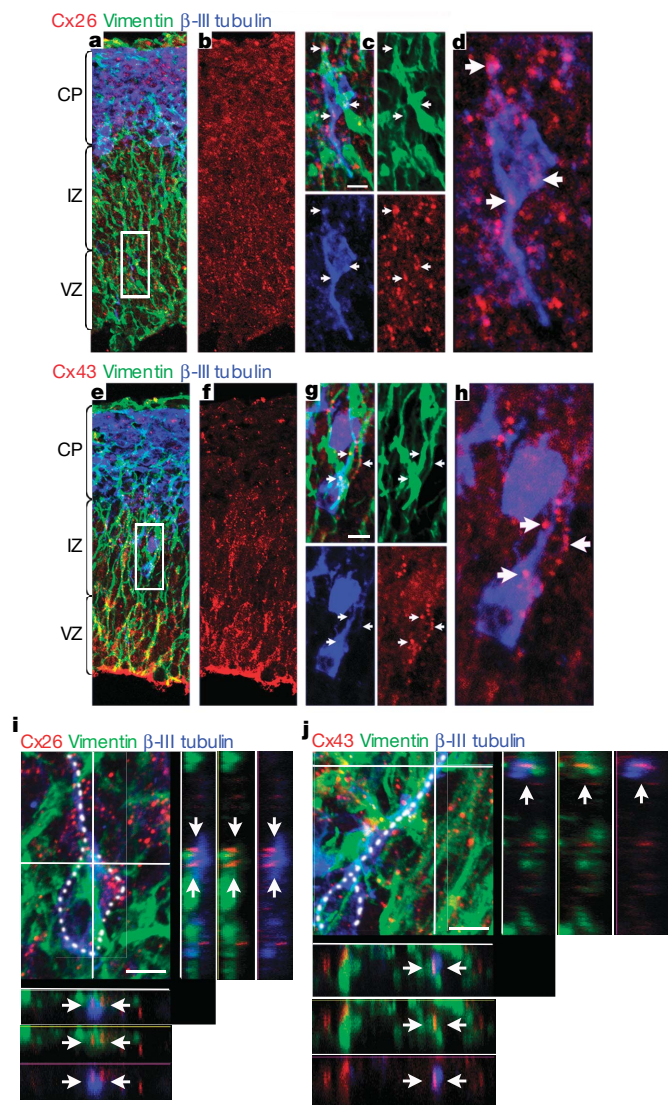


Figure 1 | Cx26 and Cx43 are localized to contact points between migrating neurons and radial glial fibres. **a, e,** Immunohistochemistry of cortical sections at E16 showing overlay of Cx26 (**a**) or Cx43 (**e**) (red), vimentin (green) to label radial fibres, and β -III tubulin (blue) to label immature neurons. **b, f,** Red channel of panels **a** and **e** showing expression pattern of Cx26 (**b**) or Cx43 (**f**). **c, g,** Migrating neuron from boxed area of panels **a** and **e**; clockwise: overlay, vimentin (green), Cx26 (**c**) or Cx43 (**g**) (red), and β -III tubulin (blue). Cx26 and Cx43 are expressed in neurons and radial glia (white arrows). **d, h,** Overlay of Cx26 (**d**) or Cx43 (**h**) and β -III tubulin in migrating neuron (from panels **c** and **g**). **i, j,** 'High resolution' confocal stacks highlighting the relationship between Cx26 (**i**) or Cx43 (**j**) puncta, vimentin⁺ fibres and a β -III tubulin⁺ neuron (dotted outline). Cross sections through the x and y axis, bottom and right panels, respectively, show individual connexin puncta (white arrows) at the interface between the vimentin⁺ radial fibre and β -III tubulin⁺ migrating neuron (Supplementary Movies 1 and 2 provide three-dimensional rotations). Scale bars, 5 μ m. CP, cortical plate; IZ, intermediate zone; VZ, ventricular zone.

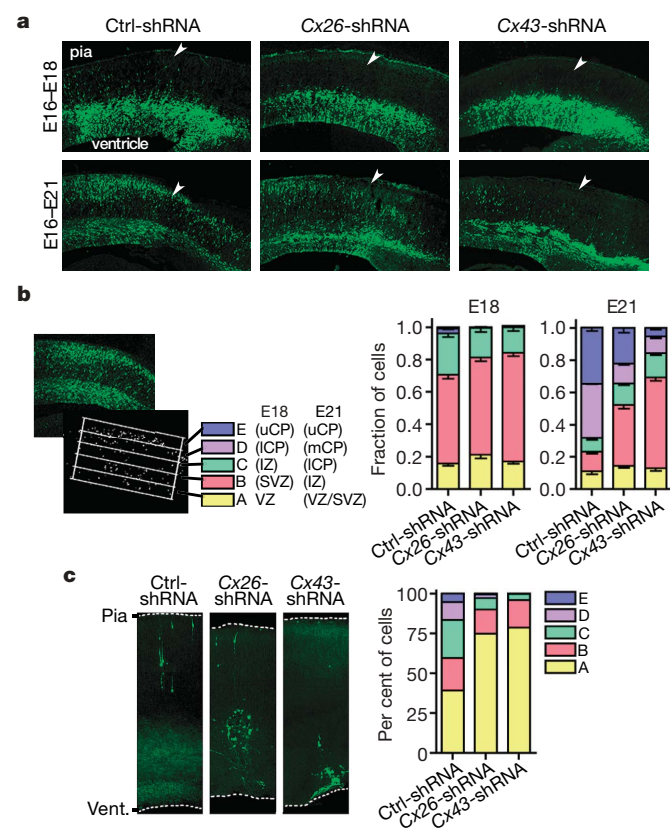


Figure 2 | Knockdown of Cx26 or Cx43 impairs neuronal migration.

a, Coronal sections from brains electroporated at E16. Cx26-shRNA or Cx43-shRNA reduces the fraction of cells in the cortical plate (white arrows) at E18 and E21. **b,** Quantification of shRNA knockdown-mediated cell distribution effect. Left panels, representative optical section with grid overlay of five equal areas (A–E) and quantification threshold. Right panels show a significant change in the distribution of cells in all areas except A at E18 and E21 (1-way ANOVA, E18: A, $P = 0.0745$; B, $P = 0.0010$; C, $P = 0.0106$; D, $P \leq 0.0001$; E, $P < 0.0001$; E21: A, $P = 0.3625$; B, $P \leq 0.0001$; C, $P = 0.0003$; D, $P \leq 0.0001$; E, $P \leq 0.0001$; $n = 3$ brains per condition). Values represent mean \pm s.e.m. **c,** Connexin expression in neurons is necessary for migration. Ctrl-shRNA-expressing cells transplanted into a wild-type host brain are able to integrate and migrate, whereas transplanted Cx26-shRNA- or Cx43-shRNA-expressing cells have reduced migration capacity (Chi-square, Ctrl-shRNA versus connexin-shRNA: A, $P < 0.0001$; B, $P < 0.0001$; C, $P > 0.05$; D, $P < 0.0001$; E, $P < 0.0001$; Ctrl-shRNA, $n = 225$ cells; Cx26-shRNA, $n = 421$; Cx43-shRNA $n = 698$). The pia surface (pia) and ventricle (vent.) are labelled to indicate orientation.

electroporated at E16 and shRNA-expressing (GFP⁺) cells were microdissected and intraventricularly injected into wild-type brains at E17. In contrast to donor Ctrl-shRNA expressing cells, which were able to engraft and migrate towards the upper layers of the cortex, the Cx26-shRNA- and Cx43-shRNA-expressing donor cells were able to engraft into the host brain but had a significantly reduced migration capacity (Fig. 2c). These data suggest that connexin expression in neurons is required for migration.

We next examined secondary defects potentially able to affect cell distribution independent of migration. Neurons expressing Cx-shRNA were able to exit the cell cycle and begin to differentiate normally (Supplementary Fig. 6a, b). Furthermore, terminal transferase dUTP nick end labelling (TUNEL) staining showed no change in apoptosis (Supplementary Fig. 6c). Additionally, the expression pattern of N-cadherin, zona occludens-1 and β_1 -integrin seemed unchanged, suggesting that connexin knockdown does not overtly disrupt expression of other cell–cell adhesion proteins (Supplementary Figs 7, 8). Together these data suggest that the defect in migration is not secondary to a defect in cell cycle exit, differentiation, cell death, or downregulation of other adhesion proteins.

Adhesions, not channels, mediate migration

What is the mechanism by which gap junctions mediate neuronal migration? We sought to differentiate among three plausible functions. Gap junctions can electrically and chemically couple cells to allow the exchange of current or small molecules, mediate extracellular release of substrates such as ATP through hemichannels, or provide an adhesive contact between two cells. We performed selective rescue experiments to distinguish these possibilities.

First, to rescue the migration phenotype, conservative mutations (CM) were introduced into the shRNA-targeted regions of Cx26 or Cx43 (Cx26CM or Cx43CM respectively), which prevented

shRNA-induced knockdown (Supplementary Fig. 9). These constructs, or enhanced yellow fluorescent protein (EYFP) alone, were electroporated with the shRNA constructs at a set molar ratio. The co-expression of Cx43CM or Cx26CM with its respective shRNA construct significantly rescued the fraction of cells that migrate to the cortical plate, demonstrating the target-specificity of the shRNA effect (Fig. 3).

To determine whether functional channels are necessary to rescue the migration defect, a conserved tyrosine in the third transmembrane domain of Cx26 and Cx43 was mutated (Cx26CMT135A or Cx43CMT154A, respectively). These dominant-negative connexin mutants are able to make adhesions but have a closed channel and thus are not able to mediate exchange between cells or with the extracellular environment²⁵ (Supplementary Figs 10, 11). Remarkably, these mutants rescued the migration of cells to the cortical plate (Fig. 3). The ability of the closed channel mutants to rescue the migration defect strongly suggests that the channel (either hemichannel or gap junction channel) does not contribute to the role of gap junctions in neuronal migration.

It has been shown that Ca²⁺ is important for neuronal migration²⁶ and that gap junctions mediate Ca²⁺ waves in the developing cortex by releasing ATP through hemichannels that activates purinergic P2Y1 receptors on radial glia²³. Thus, to further test the possibility that hemichannel-mediated Ca²⁺ waves have a role in neuronal migration, Ca²⁺ waves were inhibited by pharmacologically antagonizing the P2Y1 receptor with suramin or by knocking down the P2Y1 receptor with multiple shRNA constructs. No migration defect was observed, suggesting that hemichannel-mediated Ca²⁺ waves are not involved in cortical neuronal migration (Supplementary Fig. 12).

To determine whether the adhesive properties of gap junctions are necessary for migration, one of the conserved extracellular cysteines

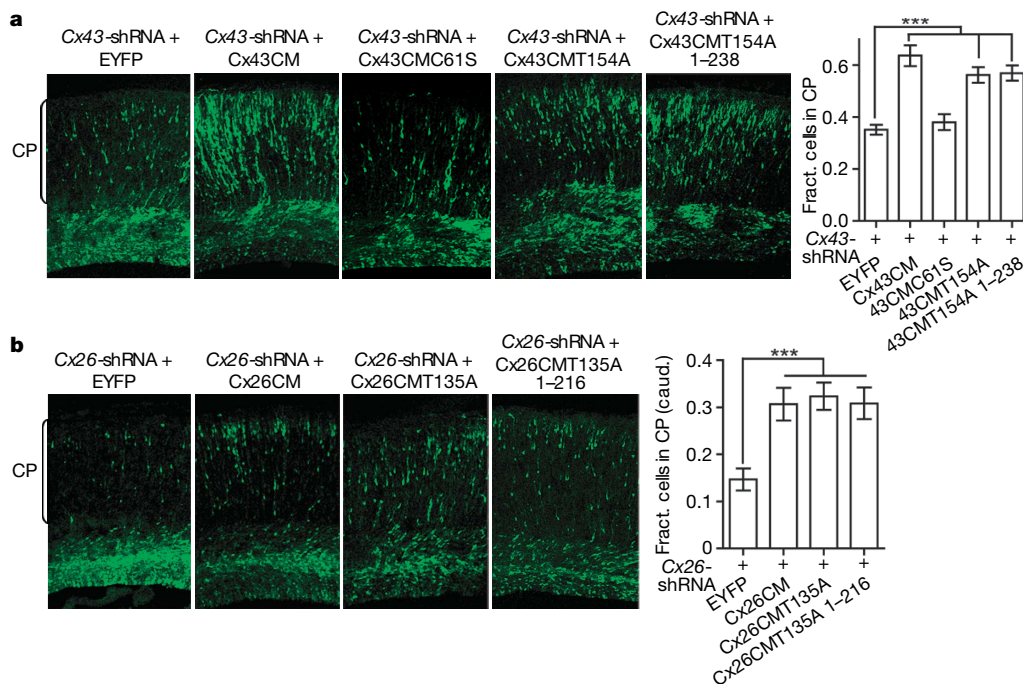


Figure 3 | Rescue of connexin-shRNA induced migration phenotype by wild-type connexin and connexin mutants that make adhesions but not channels. **a**, Cx43CM (function: channel, adhesion, C terminus), Cx43CMT154A (function: adhesion, C terminus), or Cx43CMT154A 1–238 (function: adhesion) significantly increased the fraction of cells in the cortical plate compared to EYFP when co-expressed with Cx43-shRNA ($***P < 0.0001$, *t*-test; EYFP, $n = 6$ brains; Cx43CM, $n = 4$; Cx43CMT154A, $n = 5$; Cx43CMT154A 1–238, $n = 6$). Cx43CMC61S (function: hemichannel, C terminus) did not rescue the migration defect (EYFP versus

Cx43CMC61S, *t*-test $P = 0.4727$; Cx43C61S, $n = 4$ brains). **b**, Cx26CM (function: channel, adhesion, C terminus), Cx26CMT135A (function: adhesion, C terminus), or Cx26CMT135A 1–216 (function: adhesion) significantly increased the fraction of cells in the cortical plate compared to EYFP when co-expressed with Cx26-shRNA ($***P \leq 0.0001$, *t*-test; EYFP, $n = 4$ brains; Cx26CM, $n = 5$; Cx26CMT135A, $n = 5$; Cx26CMT135A, 1–216, $n = 5$). For 26-shRNA rescue, only caudal (caud.) sections (as described in Methods) were used, because the migration phenotype is stronger caudally (Supplementary Fig. 3). Values represent means \pm s.e.m.

was mutated (Cx43CMC61S) to produce a Cx43 channel that does not make adhesions and thus cannot form gap junctions¹⁷, but maintains the ability to make functional hemichannels (Supplementary Figs 10, 11). Interestingly, Cx43CMC61S failed to rescue the Cx43-shRNA-induced migration defect, suggesting that the adhesive properties of gap junctions are necessary for radial migration (Fig. 3a).

In addition, we wanted to determine whether signalling by carboxy-terminal interactions with cytosolic proteins are important. Cx26 has a short, uncharacterized intracellular C terminus, whereas Cx43 has a longer C terminus, which interacts with a variety of proteins including zona occludens-1, V-Src, and tubulin^{27–29}. Both C-terminal truncations of the closed channel mutants (Cx43CMT154A 1–238 and Cx26CMT135A 1–216) were able to rescue the migration defect (Fig. 3). These results suggest that signalling mediated by the cytoplasmic C termini does not have a significant role during migration. However, the possibility remains that intermolecular interactions with other parts of the connexin protein may be important. In summary, we found that adhesion, but not the channel or the C terminus, is necessary for the role of gap junctions during neuronal migration.

Gap junctions provide adhesion

To demonstrate that gap junctions can mediate adhesion in cortical cells, we built on previous work showing that overexpression of Cx43, but not Cx43C61S, increases adhesion in C6 glioma cells¹⁷—a cell type without endogenous gap junction expression³⁰. We used C6 cells with or without stable transfection of Cx43–EYFP³¹ to provide a glial substrate on which we could place embryonic cortical cells and test their ability to adhere. Cortical cells were electroporated with Ctrl-shRNA, dissected and FACS-sorted for GFP, labelled with DiD, and plated on C6 cells or C6 cells expressing Cx43–EYFP (C6 + Cx43). We found a twofold increase in the number of DiD⁺ cells adhering to the C6 + Cx43 substrate compared with the C6 substrate, suggesting that cortical cells can use endogenous Cx43 as a means of cell–cell adhesion (Fig. 4a).

To test whether Cx43-shRNA reduces the ability of cortical cells to adhere to the C6 + Cx43 substrate, Ctrl-shRNA or Cx43-shRNA FACS-sorted cells were plated on C6 cells or C6 + Cx43 cells. There was a significant reduction in the ratio of the number of Cx43-shRNA to Ctrl-shRNA cells on the C6 + Cx43 substrate compared with the C6 substrate (Fig. 4b). Thus, we show that cortical cells can use endogenous Cx43 to adhere to a glial substrate and that knocking down Cx43 reduces their ability to adhere to a Cx43-expressing substrate, but not a control substrate.

Classical adhesion molecules typically provide a link with internal cytoskeletal components. Previous work in neural crest cells demonstrates that Cx43 colocalizes and co-immunoprecipitates with numerous actin-binding proteins, and that filamentous actin is disorganized in neural crest cells from the Cx43 knockout mouse³². To explore connexin–actin interactions, actin tagged with the red fluorescent protein cherry (actin–cherry) was co-electroporated with GFP. Electroporated actin is diffusely distributed in radial glia and newborn neurons and becomes punctate in multipolar SVZ cells and migrating neurons (Supplementary Fig. 13). Co-electroporation of Cx26T135A or Cx43T154A (to avoid toxicity associated with overexpression of wild-type connexin; Supplementary Fig. 11) with actin–cherry revealed that both Cx26 and Cx43 puncta frequently colocalize with the actin puncta, with the connexin puncta slightly offset (Fig. 4c). Interestingly, cells expressing Cx26-shRNA or Cx43-shRNA, have significantly fewer actin puncta than Ctrl-shRNA-expressing cells (Fig. 4d). This suggests that both Cx26 and Cx43 can interact with the actin cytoskeleton.

Connexin dynamics during migration

To investigate the dynamic behaviour of migrating neurons, we performed time-lapse live imaging. We observed that control migrating neurons typically start with a bifurcated leading process, one branch

of which prevails, and the neuron's soma then translocates forward into a dilation that forms in this leading process³³ (Fig. 5a, Supplementary Movie 3). Interestingly, migrating neurons with multiple leading processes can show differential expression of Cx43 or Cx26 in these processes (Fig. 5b). Hence, we hypothesized that gap junction adhesions play a part in stabilizing the leading process along a radial glial fibre. Indeed, Cx43-shRNA- and Cx26-shRNA-expressing neurons are unable to stabilize their processes and continue to extend multiple branches (Fig. 5c, Supplementary Movie 4, 5). This phenotype is reminiscent of neural crest cells from Cx43KO mice that show increased protrusive activity but decreased directional migration³².

To analyse the dynamics of connexin puncta localization during neuronal migration, the Cx43T154A–EYFP or the Cx26T135A–EYFP plasmid was co-electroporated with a plasmid expressing the red fluorescent protein tomato to visualize the cells, and followed by time-lapse microscopy. The closed pore connexin mutants were expressed in radial glia and migrating neurons in a pattern that resembled wild-type expression, but avoided the toxicity associated with wild-type connexin overexpression (Supplementary Fig. 11). Furthermore, because these mutants were sufficient to rescue migration, we can assume that they behave similarly to wild-type connexins with respect to neuronal migration.

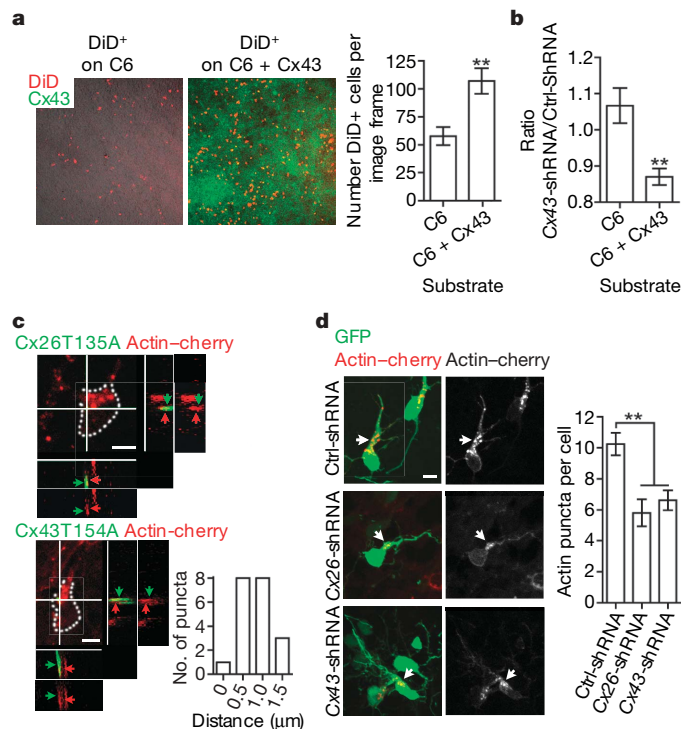


Figure 4 | Gap junctions promote cortical cell adhesion and interact with the internal actin cytoskeleton. **a**, Control DiD-labelled (red), FACS-sorted cortical cells were allowed to adhere to a glial substrate of confluent C6 or C6 + Cx43 cells. Right panel, quantification of the number of adherent DiD⁺ cells per image frame (** $P \leq 0.001$, t -test, $n = 3$ experiments). **b**, DiD-labelled, FACS-sorted Ctrl-shRNA or Cx43-shRNA expressing cells were plated on a C6 or C6 + Cx43 glial substrate. The ratio of Cx43-shRNA to Ctrl-shRNA cells that adhered to the C6 + Cx43 substrate was about 20% less than to the C6 substrate (** $P < 0.001$, t -test; C6, $n = 3$ experiments; C6 + Cx43, $n = 4$ experiments). Values represent mean \pm s.e.m. **c**, Cx26T135A–EYFP (top) or Cx43T154A–EYFP (bottom) puncta colocalize with actin–cherry (red) (cell bodies outlined in white), with their centre of fluorescence 0.5–1 μ m apart in the z -plane (coloured arrows). Graph shows a frequency histogram of the distance between the centre of the puncta. **d**, Fewer actin puncta are observed in cells expressing Cx26-shRNA or Cx43-shRNA than those expressing Ctrl-shRNA (** $P < 0.001$, t -test; Ctrl-shRNA, $n = 37$ cells; Cx26-shRNA, $n = 21$; Cx43-shRNA, $n = 44$). Images taken at ‘high resolution’. Values represent mean \pm s.e.m. Scale bars, 5 μ m.

The localization of Cx43T154A puncta in the branches of bifurcated leading processes was predictive of the dominant branch that would be maintained over time, and the absence of puncta was predictive of the transient branch (Fig. 5d, e, and Supplementary Movie 6). In the rare case in which a radial fibre was visible, stabilization of the branch along the fibre was observed (Fig. 5d, cell 2). When Cx26T135A was observed in the branches of bifurcated neurons, it also localized to the dominant branch (Supplementary Fig. 14a); however, Cx26T135A was not frequently present in the branches.

Cx26T135A was often localized to puncta in the cell body. Interestingly, Cx26T135A puncta in the cell body move into the dilation before the movement of the nucleus in approximately 81% of translocation events ($n = 56$) (Fig. 5f, Supplementary Movie 7 and Supplementary Fig. 14b). Cx43T154A puncta found in the cell body followed a similar pattern in approximately 57% of translocation events ($n = 24$) (Supplementary Fig. 14c). This stereotyped pattern of connexin puncta movement, which mimics the pattern of centrosome movement, indicates that gap junctions may provide the hypothesized adhesion contact associated with the centrosome in the dilation of translocating neurons^{33,34}. This hypothesis is supported by the observation that Cx26T135A and Cx43T154A puncta colocalize with centrin II, where the connexin puncta are slightly offset towards the cell membrane (Fig. 5g, and Supplementary Fig. 14d). Other patterns of

connexin dynamics in the trailing process and between radial fibres and neurons were also observed (Supplementary Fig. 14e, f).

Overall, the dynamics of connexin puncta localization and of shRNA electroporated cells suggest that both Cx43 and Cx26 have important roles in leading-process stabilization and somatic translocation, with Cx26 playing a more prominent part at the soma and Cx43 in the branches (Fig. 5h). Such a functional separation is consistent with the observation that Cx26 and Cx43 do not make heteromeric gap junctions, and localize to separate domains in gap junction plaques²⁴ (Supplementary Fig. 1).

Discussion

This study indicates that gap junctions mediate glial-guided radial migration of newborn neurons in the cortex. A recent study using conditional genetic deletion of Cx43 indicates that gap junctions may, in addition, mediate radial migration in the hippocampus and cerebellum³⁵. We find that gap junctions can promote adhesion in cortical cells, and that their adhesive but not their channel properties are necessary for migration. These dynamic gap junction adhesions between migrating neurons and radial glial fibres regulate properties of neuronal migration, including branch stabilization and nuclear translocation. The migration defect described here with acute knockdown of Cx43 is more dramatic than the phenotype

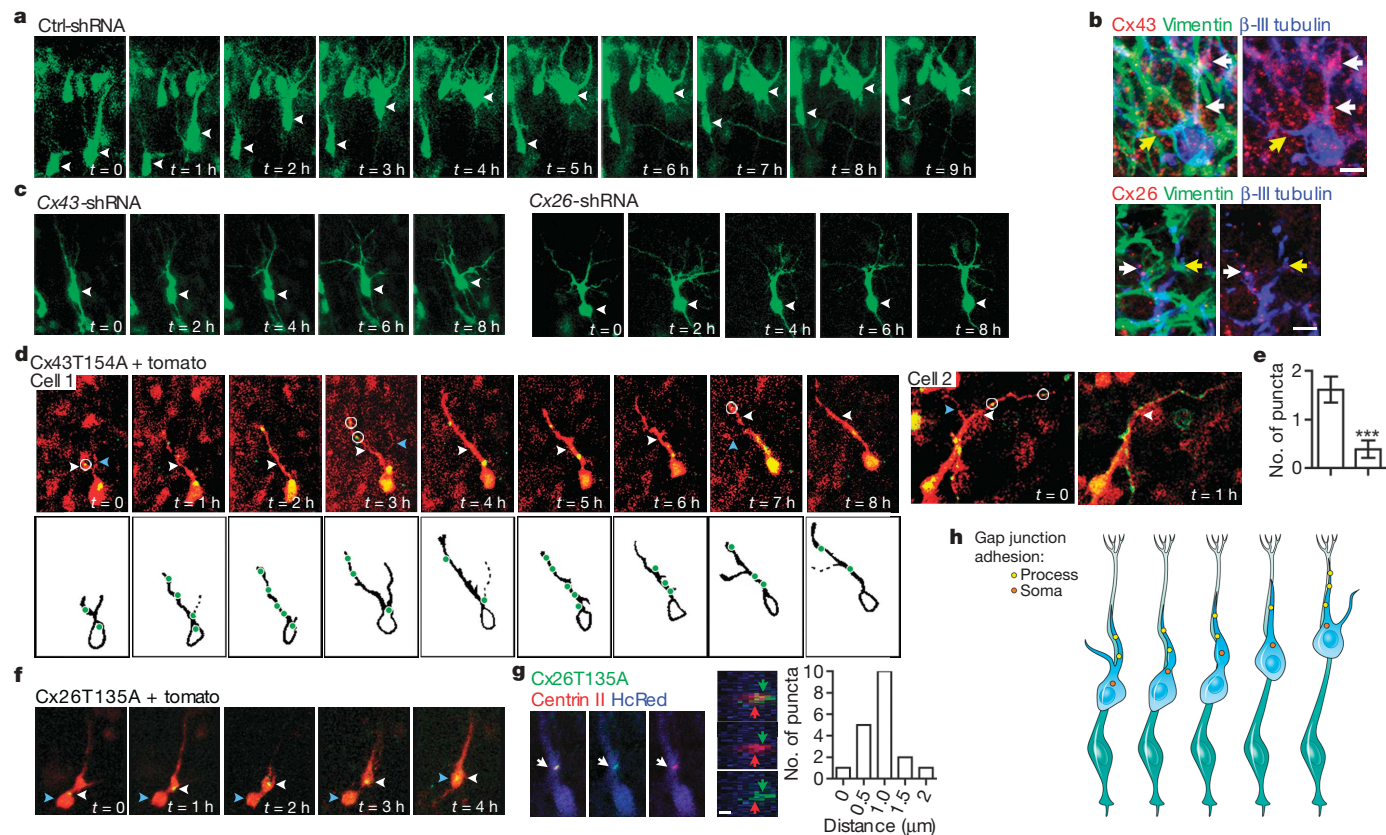


Figure 5 | Gap junction adhesions have a role in the stabilization of the migrating neuron's leading process and in the translocation of the soma. **a**, Time-lapse of two control migrating neurons (white arrowheads), demonstrating the sequence through which bifurcated neurons ($t = 0$ – 1 h) select a dominant leading process ($t = 1$ – 2 h) and translocate their soma. **b**, Immunohistochemistry for Cx43 (top) or Cx26 (bottom) in a β -III tubulin⁺ bifurcated neuron or neuronal process (relative connexin expression, white arrows > yellow arrow). Scale bars, 5 μ m. **c**, Time-lapses of neurons expressing Cx43-shRNA (left) or Cx26-shRNA (right) (see Supplementary Movies 4 and 5 for complete time-lapse). Neurons fail to stabilize a dominant process (note multiple transient processes), and the Cx26-shRNA neuron fails to translocate its soma. **d**, Time-lapses of Cx43T154A puncta in branches of bifurcated neurons. The dominant process (white arrowheads) contains numerous connexin puncta (circled in white

when multiple branches), whereas transient processes (blue arrowheads) do not. **e**, The number of puncta in the dominant (Dom.) and transient (Trans.) processes ($***P \leq 0.0001$, paired t -test, $n = 13$). Values represent mean \pm s.e.m. **f**, Time-lapse of a Cx26T135A punctum in the cell body of a migrating neuron. The punctum (white arrowhead) moves into the dilation in the leading process before the translocation of the cell body (blue arrowhead). **g**, Cx26T135A and centrin II–dsRed colocalize (white arrows) in the dilation of a migrating neuron (left panels), with their centres of fluorescence approximately 1 μ m apart in the z -plane (middle panels, coloured arrows). Frequency histogram of the distance between the centres of the puncta (right panel). Scale bar, 1 μ m. **h**, Model of gap junction adhesion dynamics in migrating neurons. Gap junction adhesive contacts stabilize the dominant leading process along radial fibres and provide an anchor in the dilation of the leading process before translocation of the soma/nucleus.

described in the *Cx43*KO mouse¹⁴. This may be due to developmental compensation in the genetic knockout arising from redundancy among connexins. Indeed, we show at least one other connexin, *Cx26*, has a very similar role in neuronal migration. Because only *Cx43* has been previously implicated in cell migration, the role of *Cx26* is interesting and implies that other connexins, including $\beta 1$ family connexins, may be involved in migration. In addition, our work indicates that different connexins may have distinct functional roles in neural migration, with *Cx26* playing a dominant role in nuclear translocation and *Cx43* in branch stabilization. Furthermore, because *Cx43* associates with other adhesion molecules including N-cadherin and integrin^{32,36,37}, it will be of interest to explore their signalling relationships. For instance, the low expression of *Cx43* in the cortical plate indicates a possible interaction with integrins—which play a role in the cessation of migration^{38,39}—and may aid in the detachment of neurons from radial fibres.

To our knowledge, a functional role for gap junctions based on adhesion, rather than channel activity, has not been described to date. Might gap junction adhesions have a role in other gap-junction-mediated cellular processes? It is thought that gap junction channels as well as their interactions with molecules such as p120catenin, integrin and the actin cytoskeleton are important for neural crest cell migration^{32,36}, and that glioblastoma invasion of the brain parenchyma requires functional gap junctions between tumour cells and astrocytes¹⁷. The migration of lung and skin cancer cells has also been associated with gap junction expression although no mechanism has been proposed^{40,41}. In addition, the neurological disorder Charcot-Marie-Tooth syndrome, caused by mutations in *Cx32*, may not only result from impaired gap junction communication but also from the loss of gap junction adhesions stabilizing Schmidt-Lanterman incisures. Further studies may elucidate an important role for the adhesive properties of gap junctions in the migration of neural crest cells and cancer cells, as well as in other diseases and physiological processes associated with gap junction function.

METHODS SUMMARY

Immunohistochemistry was performed following transcardial perfusion and cryostat sectioning. Light fixing techniques and citrate antigen retrieval were used for all connexin stains. Standard confocal images were acquired on an Olympus Fluoview 300 with 1 μm steps, and 'high resolution' images were acquired on an Olympus Fluoview 1000 with 0.5 μm steps. shRNA hairpins were inserted into the pLLOx3.7 construct for expression⁴². Plasmids were introduced into the *in-vivo* developing cortex by intraventricular injection of 1 μl of plasmid followed by electroporation⁴³. The concentration of the shRNA plasmids was 1.7 $\mu\text{g} \mu\text{l}^{-1}$, and the concentration of the rescue constructs was varied according to the specified molar ratio. Intraventricular injections were carried out in E16 timed-pregnant Sprague Dawley rats. For cell transplants, electroporations were performed at E16, GFP⁺ cells were dissected from electroporated cortices at E17 and dissociated, 50,000 cells per brain were intraventricularly injected into wild-type E17 embryos, and embryos were fixed at E21. Rat *Cx26* and *Cx43* in the pEYFPN1 vector (Clontech) were altered with QuickChange site-directed mutagenesis (Stratagene) to make the various connexin rescue constructs. For adhesion assays, electroporations were performed at E16 and cells were dissected at E19, FACS-sorted for GFP, labelled with Vybrant DiD (Molecular Probes), and plated on 100% confluent C6 cells (containing the TVA receptor) or C6 + *Cx43* cells (containing the TVA receptor and *Cx43*-EYFP)³¹, incubated for 30 min on a rotational shaker, and fixed. For actin or centrin co-expression experiments brains were electroporated at E16 and fixed at E19. For time-lapse imaging, electroporations were performed at E16, organotypic slice cultures were prepared⁴ at E18 and allowed to equilibrate overnight before images were collected using an Olympus Fluoview 300.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 April; accepted 6 July 2007.

1. Noctor, S. C., Flint, A. C., Weissman, T. A., Dammerman, R. S. & Kriegstein, A. R. Neurons derived from radial glial cells establish radial units in neocortex. *Nature* **409**, 714–720 (2001).

2. Malatesta, P., Hartfuss, E. & Gotz, M. Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neuronal lineage. *Development* **127**, 5253–5263 (2000).
3. Rakic, P. Guidance of neurons migrating to the fetal monkey neocortex. *Brain Res.* **33**, 471–476 (1971).
4. Rakic, P. Mode of cell migration to the superficial layers of fetal monkey neocortex. *J. Comp. Neurol.* **145**, 61–83 (1972).
5. Rakic, P. Specification of cerebral cortical areas. *Science* **241**, 170–176 (1988).
6. Gregory, W. A., Edmondson, J. C., Hatten, M. E. & Mason, C. A. Cytology and neuron–glial apposition of migrating cerebellar granule cells *in vitro*. *J. Neurosci.* **8**, 1728–1738 (1988).
7. Gadisseux, J. F., Kadhim, H. J., van den Bosch de Aguilar, P., Caviness, V. S. & Evrard, P. Neuron migration within the radial glial fiber system of the developing murine cerebellum: an electron microscopic autoradiographic analysis. *Brain Res. Dev. Brain Res.* **52**, 39–56 (1990).
8. Cameron, R. S. & Rakic, P. Identification of membrane proteins that comprise the plasmalemmal junction between migrating neurons and radial glial cells. *J. Neurosci.* **14**, 3139–3155 (1994).
9. Fishell, G. & Hatten, M. E. Astroactin provides a receptor system for CNS neuronal migration. *Development* **113**, 755–765 (1991).
10. Adams, N. C., Tomoda, T., Cooper, M., Dietz, G. & Hatten, M. E. Mice that lack astroactin have slowed neuronal migration. *Development* **129**, 965–972 (2002).
11. Anton, E. S., Marchionni, M. A., Lee, K. F. & Rakic, P. Role of GGF/neuregulin signaling in interactions between migrating neurons and radial glia in the developing cerebral cortex. *Development* **124**, 3501–3510 (1997).
12. Anton, E. S., Kreidberg, J. A. & Rakic, P. Distinct functions of α_3 and α_v integrin receptors in neuronal migration and laminar organization of the cerebral cortex. *Neuron* **22**, 277–289 (1999).
13. Nadarajah, B., Jones, A. M., Evans, W. H. & Parnavelas, J. G. Differential expression of connexins during neocortical development and neuronal circuit formation. *J. Neurosci.* **17**, 3096–3111 (1997).
14. Fushiki, S. *et al.* Changes in neuronal migration in neocortex of connexin43 null mutant mice. *J. Neuropathol. Exp. Neurol.* **62**, 304–314 (2003).
15. Huang, G. Y. *et al.* Gap junction-mediated cell–cell communication modulates mouse neural crest migration. *J. Cell Biol.* **143**, 1725–1734 (1998).
16. Lo, C. W., Waldo, K. L. & Kirby, M. L. Gap junction communication and the modulation of cardiac neural crest cells. *Trends Cardiovasc. Med.* **9**, 63–69 (1999).
17. Lin, J. H. *et al.* Connexin 43 enhances the adhesivity and mediates the invasion of malignant glioma cells. *J. Neurosci.* **22**, 4302–4311 (2002).
18. Oliveira, R. *et al.* Contribution of gap junctional communication between tumor cells and astroglia to the invasion of the brain parenchyma by human glioblastomas. *BMC Cell Biol.* **6**, 7 (2005).
19. Harris, A. L. Emerging issues of connexin channels: biophysics fills the gap. *Q. Rev. Biophys.* **34**, 325–472 (2001).
20. Dermietzel, R. *et al.* Differential expression of three gap junction proteins in developing and mature brain tissues. *Proc. Natl Acad. Sci. USA* **86**, 10148–10152 (1989).
21. Lo Turco, J. J. & Kriegstein, A. R. Clusters of coupled neuroblasts in embryonic neocortex. *Science* **252**, 563–566 (1991).
22. Bittman, K., Owens, D. F., Kriegstein, A. R. & LoTurco, J. J. Cell coupling and uncoupling in the ventricular zone of developing neocortex. *J. Neurosci.* **17**, 7037–7044 (1997).
23. Weissman, T. A., Riquelme, P. A., Ivic, L., Flint, A. C. & Kriegstein, A. R. Calcium waves propagate through radial glial cells and modulate proliferation in the developing neocortex. *Neuron* **43**, 647–661 (2004).
24. Falk, M. M. Connexin-specific distribution within gap junctions revealed in living cells. *J. Cell Sci.* **113**, 4109–4120 (2000).
25. Beahm, D. L. *et al.* Mutation of a conserved threonine in the third transmembrane helix of α - and β -connexins creates a dominant-negative closed gap junction channel. *J. Biol. Chem.* **281**, 7994–8009 (2006).
26. Komuro, H. & Rakic, P. Intracellular Ca^{2+} fluctuations modulate the rate of neuronal migration. *Neuron* **17**, 275–285 (1996).
27. Giepmans, B. N. & Moolenaar, W. H. The gap junction protein connexin43 interacts with the second PDZ domain of the zona occludens-1 protein. *Curr. Biol.* **8**, 931–934 (1998).
28. Lin, R., Warn-Cramer, B. J., Kurata, W. E. & Lau, A. F. v-Src phosphorylation of connexin 43 on Tyr247 and Tyr265 disrupts gap junctional communication. *J. Cell Biol.* **154**, 815–827 (2001).
29. Giepmans, B. N. *et al.* Gap junction protein connexin-43 interacts directly with microtubules. *Curr. Biol.* **11**, 1364–1368 (2001).
30. Naus, C. C., Bechberger, J. F., Caveney, S. & Wilson, J. X. Expression of gap junction genes in astrocytes and C6 glioma cells. *Neurosci. Lett.* **126**, 33–36 (1991).
31. Lai, A. *et al.* Oculodentodigital dysplasia connexin43 mutations result in non-functional connexin hemichannels and gap junctions in C6 glioma cells. *J. Cell Sci.* **119**, 532–541 (2006).
32. Xu, X., Francis, R., Wei, C. J., Linask, K. L. & Lo, C. W. Connexin 43-mediated modulation of polarized cell movement and the directional migration of cardiac neural crest cells. *Development* **133**, 3629–3639 (2006).
33. Schaar, B. T. & McConnell, S. K. Cytoskeletal coordination during neuronal migration. *Proc. Natl Acad. Sci. USA* **102**, 13652–13657 (2005).

34. Tsai, J. W., Bremner, K. H. & Vallee, R. B. Dual subcellular roles for LIS1 and dynein in radial neuronal migration in live brain tissue. *Nature Neurosci.* **10**, 970–979; advance online publication, doi:10.1038/nn1934 (8 July 2007).
35. Wiencken-Barger, A. E., Djukic, B., Casper, K. B. & McCarthy, K. D. A role for Connexin43 during neurodevelopment. *Glia* **55**, 675–686 (2007).
36. Xu, X. *et al.* Modulation of mouse neural crest cell motility by N-cadherin and connexin 43 gap junctions. *J. Cell Biol.* **154**, 217–230 (2001).
37. Meyer, R. A., Laird, D. W., Revel, J. P. & Johnson, R. G. Inhibition of gap junction and adherens junction assembly by connexin and A-CAM antibodies. *J. Cell Biol.* **119**, 179–189 (1992).
38. Dulabon, L. *et al.* Reelin binds $\alpha 3 \beta 1$ integrin and inhibits neuronal migration. *Neuron* **27**, 33–44 (2000).
39. Schmid, R. S., Jo, R., Shelton, S., Kreidberg, J. A. & Anton, E. S. Reelin, integrin and DAB1 interactions during embryonic cerebral cortical development. *Cereb. Cortex* **15**, 1632–1636 (2005).
40. el-Sabban, M. E. & Pauli, B. U. Adhesion-mediated gap junctional communication between lung-metastatic cancer cells and endothelium. *Invasion Metastasis* **14**, 164–176 (1994).
41. Ito, A. *et al.* A role for heterologous gap junctions between melanoma and endothelial cells in metastasis. *J. Clin. Invest.* **105**, 1189–1197 (2000).
42. Lois, C., Hong, E. J., Pease, S., Brown, E. J. & Baltimore, D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295**, 868–872 (2002).
43. Saito, T. & Nakatsuji, N. Efficient gene transfer into the embryonic mouse brain using *in vivo* electroporation. *Dev. Biol.* **240**, 237–246 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful for ideas arising from discussions with G.M. Elias, members of the Kriegstein Laboratory, A. Alvarez-Buylla, J. L. Rubenstein and S. J. Pleasure, as well as manuscript and figure editing by G.M. Elias. We thank D. Laird for connexin plasmids, A. Lai for C6 cell lines, K. Hu for the actin-cherry plasmid, R. Vallee for the centrin II-dsRed plasmid, and W. Walantus, J. Agudelo and T. Calcagni for technical support. Artwork in Fig. 5h is by K. X. Probst. This work was supported by grants from the National Institutes of Health (to A.R.K.), the Sandler Family and Genentech Graduate Fellowship (to L.A.B.E.), the California Institute for Regenerative Medicine Graduate Fellowship (to L.A.B.E.), and the J.G. Bowes Research Fund.

Author Contributions L.A.B.E. conceived of and carried out all experiments except as noted below. D.D.W. developed methods for and carried out the cell transplant/cell autonomy experiments and the whole-cell patch clamp recordings in C6 cells. A.R.K., as the principle investigator, provided conceptual and technical guidance for all aspects of the project. L.A.B.E. wrote the manuscript. All authors discussed the results/experiments and revised/edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.A.B.E. (EliasL@stemcell.ucsf.edu) or A.R.K. (KriegsteinA@stemcell.ucsf.edu).

METHODS

Immunohistochemistry and imaging. Embryos were fixed by transcardial perfusion with PBS followed by 4% paraformaldehyde in PBS (PFA). A subsequent perfusion of PBS was performed in embryos used for Cx26/43 staining. Brains were post-fixed overnight (except for Cx26/43 staining), dehydrated in 20% and 30% sucrose, frozen, and cryostat sectioned at 14 μm (except for Cx26/43 staining, which used 10 μm). For endogenous knockdown assays, dissociated cortical cells were plated on Poly-L-Lysine coated coverslips, allowed to adhere for 2 h, and fixed in PFA (10 min). Citrate antigen retrieval was performed for Cx26/43 and Ki67 staining. Blocking solution contained 2% gelatin, 10% serum and 1% Triton-X-100 in PBS. Sections/cells were incubated overnight in primary antibodies: rabbit anti-Cx43 (1:50, Zymed), mouse anti-Cx43 (1:50, Zymed), rabbit anti-Cx26 (1:25, Zymed SI-2800), chicken anti-vimentin (1:1,000, Chemicon), mouse anti- β -III tubulin (TUBJ1) (1:200, Covance), chicken anti-GFP (1:500, Aves Labs), rabbit anti-Ki67 (1:1,000, Novocastra), mouse anti-bromodeoxyuridine Alexa Fluor 647 (1:50, Molecular Probes), mouse anti-nestin (1:100, Chemicon), rabbit anti ZO-1 (1:100, Zymed), mouse anti N-cadherin (1:100, Zymed), and mouse anti β 1-integrin (1:100, BD Transduction Laboratories). TUNEL staining was performed using TMR red Cell Death Detection Kit (Roche). Secondary antibodies included Alexa 488/568 conjugates (Molecular Probes) or Cy3/Cy5 conjugates (Jackson). Standard images were acquired on an Olympus Fluoview 300 laser-scanning confocal microscope. 'Standard' images of connexin staining represent collapsed $\leq 3 \mu\text{m}$ stacks collected at 1- μm steps, and other standard images represent single optical sections. 'High resolution' images were collected on an Olympus Fluoview 1000 at 0.5- μm steps. Images were analysed using Photoshop 7.0.

Plasmid constructs. ShRNA oligonucleotides were inserted into the dual promoter construct pLlox3.7 for expression under the U6 promoter (EGFP was expressed under the CMV promoter)⁴². The following target sequences were used for Cx26, TCTGGAATTGTCATCCTGCTA (alternative target: GCAGC-GTCTGGTGAAGTGTA); and Cx43, GCAATTACAACAAGCAAGCTA (alternative target: GGCTTGCTGAGAACCTACATCATCA). The control shRNA oligonucleotides with three point mutations (underlined) are as follows for Cx26, TCTTGAATATGTCATCGTGCTA; and Cx43, GTAATTGCAACAAGAA-AGCTA. The following target sequences were used for the P2Y1 receptor: P2Y1-shRNA(2) GGAGTGAGGCCAATTTACA, P2Y1-shRNA(3) GAGTACCTGCG-AAGTTATT.

Rat Cx26 and Cx43 clones in the pEYFP-N1 vector (Clontech) were a gift of D. Laird. QuickChange site-directed mutagenesis (Stratagene) was used to introduce the following point mutations as underlined: Cx26CM, CTCTGTGTCGG-GAATCTGCATACTGCTCAACATCAGAGCTGTG; Cx43CM, CCTCGTG-CCGGAAATTATAACAAACAAGCCAGCGAGCAAACTGGG; Cx43C61S, CT-CAACAACCTGGCTCCGAAACGTCTGCTATGAC; Cx26T135A, TCCCTG-TGGTGGGCTACACACCAGC; Cx43T54A, GGCTTGCTGAGAGCCTAC-ATCATCAGCATCC; Cx26 1-216, ATCAGAGCTGTGCTATCTGTTTCATT-AGAATTGCTCAGGGAAGTCC; Cx43 1-238, TTCAAAGGCGTTAAGGA-TCCCGTGAAGGGAAG (introduced restriction site in bold). The P2Y1 receptor was cloned with the following primers: AGCTCGAGTGCCTGAGTTGG-AAAGAA (sense), TCGAATTCGGGCGTAGTCGGGCACGTCGTAGGGG-TACAAACTGTGTCTCCGTT (antisense), and inserted into the pEYFPN1 vector (Clontech) using the *Xho*I and *Eco*R1 restriction sites. The antisense primer introduced a haemagglutinin (HA) tag and removed the stop codon.

Cell transfections and western blots. Western blots were performed on Cos-7 cells transfected with FuGENE6 transfection reagent (Roche) with 1 μg of DNA, 200 μl Opti MEM and 5 μl FuGENE6 per 35 mm dish. Samples were collected 3 days post transfection in RIPA buffer (150 mM NaCl, 0.1% SDS, 1% Triton X100, 50 mM TrisHCl pH 6.8, 1 mM EDTA, 0.5% deoxycholate, 10% glycerol) plus protease inhibitors, normalized for protein concentration, denatured at 37 $^{\circ}\text{C}$ for 30 min in 4 \times loading buffer (6 mM Tris pH 6.8, 5% β -Mercaptoethanol, 2% SDS, 0.5% bromophenol blue), and run on Criterion Precast 12.5% Tris-HCl gels (Bio-Rad). Membranes were probed with rabbit anti-Cx43 (1:1,000, Zymed) or rabbit anti-Cx26 (1:2,000, Zymed 71-0500) followed by anti-rabbit peroxidase conjugate (1:2,000, Sigma) or with mouse anti-HA peroxidase for P2Y1-HA detection (1:2,000, Sigma H 6533). All blots were stripped and re-probed with mouse anti- γ -tubulin (1:500, Sigma) and anti-mouse peroxidase conjugate (1:400, Sigma). For quantification, all western blots were normalized to the levels

of the loading control γ -tubulin and band intensity was measured in Photoshop. C6 cells were transfected with FuGENE6 transfection reagent (Roche) with 0.3 μg of DNA, 20 μl Opti MEM and 2 μl FuGENE6 per 16 mm dish.

In-utero injection and electroporation/cell transplants. Animals were maintained according to protocols approved by the Institutional Animal Care and Use Committee at UCSF. Plasmids were introduced into the developing cortex *in vivo* by intraventricular injection and electroporation⁴³. Intraventricular injections were carried out in E16 timed-pregnant Sprague Dawley rats as previously described¹. Electroporations were performed using an Electro Square Porator ECM830 (Genetronics) (5 pulses, 50 V, 100 ms, 1-s interval). DNA was prepared in endotoxin-free conditions and 1 μl was injected per brain at the following concentrations: shRNA constructs, 1.7 $\mu\text{g} \mu\text{l}^{-1}$; rescue constructs, at a molar ratio of 0.67:1 (Cx26 rescue construct: Cx26-shRNA) or 0.45:1 (Cx43 rescue construct: Cx43-shRNA); Cx26T135A and Cx43T154A 1.7 $\mu\text{g} \mu\text{l}^{-1}$; Tomato 1 $\mu\text{g} \mu\text{l}^{-1}$; actin-cherry 1.7 $\mu\text{g} \mu\text{l}^{-1}$; centrin II-dsRed 2.0 $\mu\text{g} \mu\text{l}^{-1}$; HcRed 2.5 $\mu\text{g} \mu\text{l}^{-1}$. For analysis of endogenous knockdown, electroporations were performed at E16 and cortices were dissociated at E18, allowed to adhere in culture for 2 h, fixed and stained. For cell transplants, electroporations were performed at E16, GFP⁺ cells were dissected from electroporated cortices at E17 in ACSF (125 mM NaCl, 2.5 mM KCl, 1 mM MgCl₂, 2 mM CaCl₂, 1.25 mM NaH₂PO₄, 25 mM NaHCO₃ and 20 mM glucose), the cells were dissociated and resuspended in culture media containing 25% Hanks BBS, 47% Basal Medium Eagle (Gibco), 5% Fetal bovine serum, 1X Pen/Strep, 0.66% glucose, 1% N2-supplement (Invitrogen). Fifty-thousand cells per brain were intraven- tricularly injected into wild-type E17 embryos, and embryos were fixed at E21. The video protocol in ref. 44 fully explains the *in utero* intraventricular injections and electroporations.

Quantification of cell distribution. The shRNA cell distribution phenotype was quantified by taking a total of 20 evenly spaced rostral to caudal sections per brain from 3 brains per condition. A threshold was applied in Photoshop and the fraction of cells in each of 5 rows of equal area (A–E) stacked parallel to the ventricular zone surface was quantified in each slice and then averaged across slices. The Cx43-shRNA rescue experiments were quantified in a similar manner by counting the fraction of cells in the cortical plate (layers C, D and E) in 10 rostral to caudal sections per brain for 4–6 brains per condition. For the Cx26-shRNA rescue experiments, 5 medial to caudal sections per brain for 4–6 brains per condition were quantified. For transplant experiments, confocal stacks of all sections with cells that integrated into the cortex were collected and the number of cells in each area (A–E) was counted. All quantifications were performed in a blind manner.

Cell sorting. Electroporations were performed at E16 and cells were dissected at E19 in ACSF. Cells were transferred to PBS 2%FBS immediately before sorting and kept on ice. Cells were sorted for GFP on a FACSaria (Becton Dickinson) using a 100 μm nosel, 488 laser with the 530/30 filter, and were collected in culture media. Collected cells were labelled with Vybrant DiD cell labelling solution (Molecular Probes) in serum-free culture media for 20 min and washed according to the instructed protocol. Ten-thousand cells were plated per well of 100% confluent C6 cells grown on 24-well collagen biocoat plates (Becton Dickinson). C6 (containing the TVA receptor) and C6 + Cx43 (containing the TVA receptor and Cx43-EYFP) cells were a gift of A. Lai³¹. Cells were incubated for 30 min, with shaking at 250 rpm, and then washed 4 times with PBS, and fixed. Each experiment included three or more wells of each condition. Five pictures of each well were taken (middle, right, left, top, bottom), and quantifications were carried out in a blind manner.

Time-lapse imaging. Electroporations were performed at E16 and 300- μm thick vibratome sections were prepared at E18 in ACSF. Slices were plated on Millicell-CM inserts (Millipore) in culture medium and allowed to equilibrate overnight before imaging. Confocal images, were collected on the Olympus Fluoview 300 at 2–5- μm steps (to minimize bleaching) every hour for 8–10 h. Stacks containing the cell of interest were collapsed and analysed in Photoshop. The video protocol in ref. 45 fully explains the slice culture system used for time-lapse experiments.

44. Walantus, W., Elias, L. & Kriegstein, A. R. *In utero* intraventricular injection and electroporation of E16 rat embryos. *J. Visualized Exp.* 6 (<http://www.jove.com/index/Details.stp?ID=236&VID=222>) (2007).

45. Elias, L. & Kriegstein, A. R. Organotypic slice culture of E18 rat brains. *J. Visualized Exp.* 6 (<http://www.jove.com/index/Details.stp?ID=235&VID=221>) (2007).

LETTERS

No extreme bipolar glaciation during the main Eocene calcite compensation shift

Kirsty M. Edgar¹, Paul A. Wilson¹, Philip F. Sexton^{1†} & Yusuke Suganuma²

Major ice sheets were permanently established on Antarctica approximately 34 million years ago^{1–3}, close to the Eocene/Oligocene boundary, at the same time as a permanent deepening of the calcite compensation depth in the world's oceans⁴. Until recently, it was thought that Northern Hemisphere glaciation began much later, between 11 and 5 million years ago^{1–3,5}. This view has been challenged, however, by records of ice rafting at high northern latitudes during the Eocene epoch^{6,7} and by estimates of global ice volume that exceed the storage capacity of Antarctica⁸ at the same time as a temporary deepening of the calcite compensation depth ~41.6 million years ago⁹. Here we test the hypothesis that large ice sheets were present in both hemispheres ~41.6 million years ago using marine sediment records of oxygen and carbon isotope values and of calcium carbonate content from the equatorial Atlantic Ocean. These records allow, at most, an ice budget that can easily be accommodated on Antarctica, indicating that large ice sheets were not present in the Northern Hemisphere. The records also reveal a brief interval shortly before the temporary deepening of the calcite compensation depth during which the calcite compensation depth shoaled, ocean temperatures increased and carbon isotope values decreased in the equatorial Atlantic. The nature of these changes around 41.6 million years ago implies common links, in terms of carbon cycling, with events at the Eocene/Oligocene boundary⁴ and with the 'hyperthermals' of the Early Eocene climate optimum^{3,10,11}. Our findings help to resolve the apparent discrepancy between the geological records of Northern Hemisphere glaciation^{6–8} and model results^{12,13} that indicate that the threshold for continental glaciation was crossed earlier in the Southern Hemisphere than in the Northern Hemisphere.

A striking feature of the composite³ Cenozoic $\delta^{18}\text{O}$ record in benthic foraminiferal calcite is the abrupt increase in values across the Eocene/Oligocene boundary, 34 Myr ago (Fig. 1a). Together with evidence from Southern Ocean records for the initiation of ice rafting¹⁴ and glacial weathering, and from sequence stratigraphic studies¹⁵ for a global sea level fall (~70 m apparent sea level), this $\delta^{18}\text{O}$ shift is widely interpreted to signify the onset of major Antarctic glaciation^{1–3,14} (Fig. 1a). The standard view^{1–3,5} is that substantial Northern Hemisphere glaciation occurred much later (Fig. 1a); recently, however, three lines of evidence have challenged this orthodoxy.

First, evidence now exists for ice rafting in the Arctic Ocean around 45 Myr ago⁶ and in the Norwegian–Greenland Sea between 37 and 27 Myr ago⁷ (Fig. 1a). Second, new records for the Eocene/Oligocene boundary from the equatorial Pacific Ocean (ODP Leg 199 of the Ocean Drilling Program) show⁴ that the amplitude of the $\delta^{18}\text{O}$ change ($\Delta\delta^{18}\text{O}_{\text{benthic}} = 1.5\text{‰}$ VPDB) is impossibly large to be attributed to Antarctic glaciation alone. The simplest explanation for this result is that some of the $\Delta\delta^{18}\text{O}_{\text{benthic}}$ signal must denote global cooling⁴. However, Mg/Ca records^{2,16} provide little support for declining

temperatures. This raises the possibility of ice growth beyond Antarctica and/or the operation of some factor (such as changes in seawater carbonate chemistry associated with increasing calcite compensation depth, CCD) acting to mask the cooling signal in the Mg/Ca records^{4,16}. Third, the $\delta^{18}\text{O}$ increase associated with the Eocene/Oligocene boundary occurred in lock-step with permanent deepening of the CCD, possibly in response to glacioeustatic sea level fall and reduced shelf carbonate accumulation⁴. Thus, the discovery^{9,17} of multiple intervals of temporary CCD deepening in the equatorial Pacific earlier in the Eocene (Fig. 1b) has prompted speculation of transient glaciations before the Eocene/Oligocene boundary. A recent study⁸ invokes three such glaciations including one extreme bipolar event (100–190 m apparent sea level change; $\sim 35\text{--}70 \times 10^6 \text{ km}^3$ ice) coincident with the most prominent of the transient Eocene CCD deepening events (carbonate accumulation event CAE-3 ~41.6 Myr ago; ref. 9 and Fig. 1b, c). Yet the idea that large ice sheets existed in both hemispheres at this time is at odds with contemporaneous warm polar ocean temperatures^{2,3,18}, high atmospheric CO_2 levels¹⁹ (Fig. 1a) and the occurrence of (sub)tropical flora at mid- to high latitudes²⁰. Sequence stratigraphic records²¹ support the concept of early Cenozoic glaciations, but the ice sheets invoked are much more modest in size (~10 to 45 m apparent sea level change; ~40 to 65% of present East Antarctic ice sheet) and considered compatible with high-latitude warmth because of their restriction to the interior of Antarctica²¹.

Although the sensitivity to CCD change¹⁷ of the Pacific drill sites used to infer extreme bipolar Eocene glaciation is beneficial to understanding carbon cycling, it is detrimental to generating isotope records from calcareous microfossils whose occurrence and preservation are sensitive to changes in carbonate saturation state. Thus, further tests of bipolar Eocene glaciation are warranted. We generated high-resolution (~4.5 kyr) monospecific stable isotope records for the interval spanning CAE-3 (ref. 9) from the equatorial Atlantic Ocean (ODP Leg 207, Demerara Rise, Site 1260, 9° 16' N, 54° 33' W; palaeowater depth 2,500–3,000 m)²². These sediments were deposited well above the local CCD for most of the study interval and are shallowly buried²², favouring calcareous microfossil preservation¹⁸.

In Fig. 1c–e, we compare our new $\delta^{18}\text{O}$ records from the equatorial Atlantic with the record presented⁸ from the equatorial Pacific in support of bipolar glaciation. $\delta^{18}\text{O}$ data from the equatorial Pacific show a considerable spread, and the amplitude of increase used to infer extreme bipolar glaciation relies on outliers in sparse data ($\Delta\delta^{18}\text{O}_{\text{benthic}} = 1.2\text{‰}$ VPDB, arrows, Fig. 1c). Our data from the Atlantic are more continuous and show less spread (Fig. 1d, e). Our benthic record (Fig. 1d) shows a well-defined minimum around the middle of magnetochron C19r followed by a shift to higher values attained by the onset of CAE-3 (zone I/II boundary, Fig. 1c). Our planktic record shows a similar pattern (Fig. 1e). One interpretation of the correspondence between our planktic and benthic records is

¹National Oceanography Centre, School of Ocean and Earth Science, European Way, Southampton, SO14 3ZH, UK. ²Department of Earth & Planetary Science, University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo, 113-0033 Japan. [†]Present address: Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0244, USA.

that it reflects an increase in the oxygen isotope composition of sea water ($\delta^{18}\text{O}_{\text{seawater}}$) and supports the notion of ice growth. Another interpretation, arguably more in keeping with the other palaeoclimate records¹⁹ and numerical modelling results^{12,13} shown in Fig. 1a, is that our data are attributable to cooling of surface and bottom waters at Site 1260. Regardless, our estimate of the $\Delta\delta^{18}\text{O}_{\text{benthic}}$ signal associated with CAE-3 is about half that proposed⁸ from the Pacific (Fig. 1d versus Fig. 1c).

In Fig. 2 we show the relationship between measured $\Delta\delta^{18}\text{O}_{\text{benthic}}$ and global continental ice volumes for a range of values for mean $\delta^{18}\text{O}$ of stored ice ($\delta^{18}\text{O}_{\text{ice}}$). Horizontal dashed lines denote different values for $\Delta\delta^{18}\text{O}_{\text{benthic}}$, vertical solid lines correspond to different values assumed for $\delta^{18}\text{O}_{\text{ice}}$ and intersections yield the resulting estimated ice volumes (see 'isovol' lines). Lines a and b assume that the whole $\Delta\delta^{18}\text{O}_{\text{benthic}}$ signal is attributable to ice volume. Assuming that $\delta^{18}\text{O}_{\text{ice}}$ was as extreme as today (about -50‰ SMOW^{4,23}), the $\Delta\delta^{18}\text{O}_{\text{benthic}}$ signal inferred in the Pacific study⁸ for CAE-3 (line a in Fig. 2) yields a global middle Eocene ice budget almost 1.5 times the present Antarctic ice volume ($25.4 \times 10^6 \text{ km}^3$)²⁴. However, the latitudinal temperature gradient during the Eocene was less extreme than today, so it is likely that $\delta^{18}\text{O}_{\text{ice}}$ was also less extreme (conservatively, about -30‰ SMOW^{4,23,25}), yielding even larger ice budgets (almost 2.5 times the modern Antarctic budget). This ice volume is close to the global total estimated for the last glacial maximum (Fig. 2), when large ice sheets existed not only on Antarctica and Greenland but also over large areas of North America and Eurasia, and it implies an even greater global ice budget for the Eocene/Oligocene boundary when benthic $\delta^{18}\text{O}$ values were $\sim 0.8\text{‰}$ higher

(Fig. 1a). Even larger ice volumes were invoked⁸ for CAE-3 in the Pacific study using the dual benthic $\text{Mg}/\text{Ca} - \delta^{18}\text{O}$ method ($\Delta\delta^{18}\text{O}_{\text{seawater}} = 1.5\text{‰}$). But we focus on the $\Delta\delta^{18}\text{O}_{\text{benthic}}$ signal alone because the presence of authigenic dolomite^{16,17} towards the base of the section at ODP Site 1218 means that the only potentially meaningful Mg/Ca data published for CAE-3 are for its termination (see figure 3 of ref. 8). By focusing on the $\Delta\delta^{18}\text{O}_{\text{benthic}}$ signal alone we adopt a conservative approach because the ice budgets calculated in this way in ref. 8, although extreme (because the whole signal was attributed to ice growth with no associated cooling), are less extreme than those that were estimated by the dual benthic $\text{Mg}/\text{Ca} - \delta^{18}\text{O}$ method.

If we assume no cooling component associated with the smaller $\Delta\delta^{18}\text{O}_{\text{benthic}}$ signal seen in our new record (line b, Fig. 2), our revised estimate of the upper limit of possible ice growth associated with CAE-3 is correspondingly smaller, it need not require ice storage in the Northern Hemisphere and it is more in keeping with sequence stratigraphy records²¹. On this basis, we cannot exclude the possible existence of small valley glaciers draining the uplands of Greenland around 41.6 Myr ago, particularly in light of dropstones recently discovered^{6,7} in Eocene strata in the high northern latitudes. Nevertheless, we can rule out the existence of large ice sheets in the Northern Hemisphere and we therefore find no support for extreme Eocene bipolar glaciation. Based on work elsewhere^{23,26} we might expect the $\Delta\delta^{18}\text{O}_{\text{benthic}}$ signal to be composed of at least equal parts ice volume and temperature (line b', Fig. 2). On this basis we calculate ice budgets that are easily accommodated on central Antarctica alone (~ 0.4 to 0.6 times modern Antarctic budget). Under these circumstances, only by assuming an extremely high value for

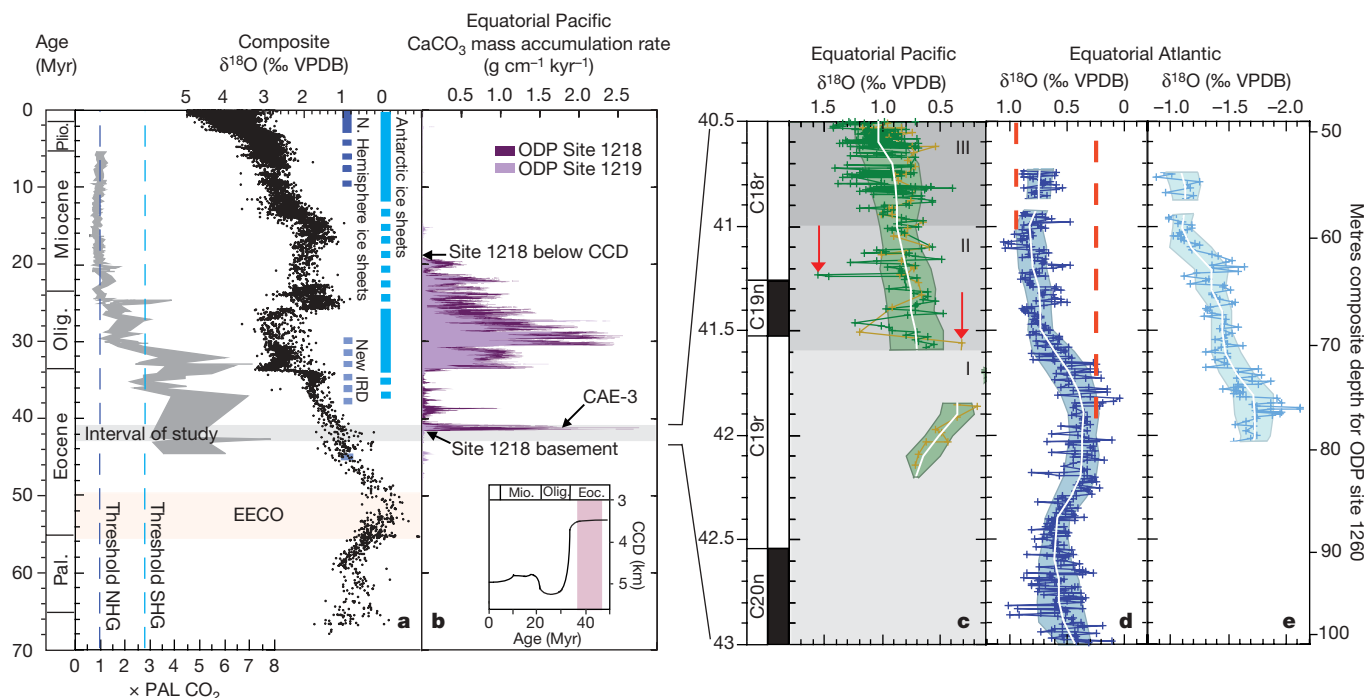
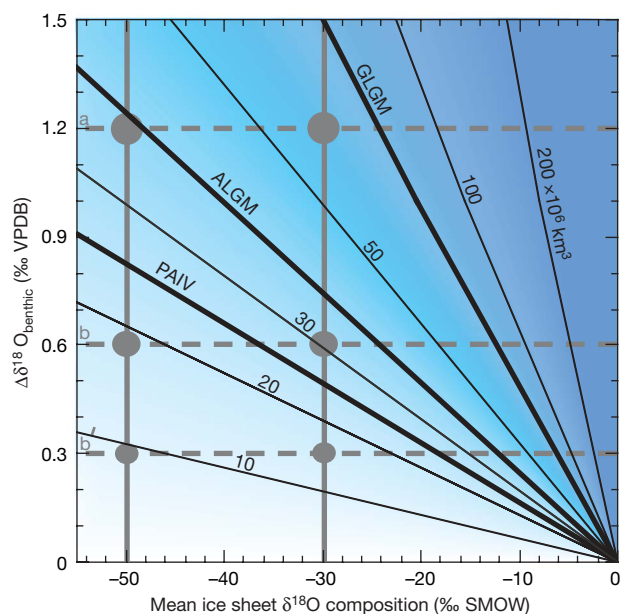


Figure 1 | Palaeoceanographic records showing changes in ocean chemistry and global climate. **a**, Cenozoic atmospheric CO_2 (ref. 19), grey shading defines the range of estimated CO_2 values as calculated in ref. 19 relative to the pre-anthropogenic level (PAL). Vertical dashed blue lines are model-defined thresholds for large ice sheet initiation. SHG and NHG, Southern and Northern Hemisphere glaciation^{12,13}. Composite benthic $\delta^{18}\text{O}$ record³. Vertical blue bars show the standard view of the presence and extent of full scale/permanent ice sheets (solid bars) and those thought to have been partial/ephemeral (broken bars). Occurrence of newly documented ice rafted debris (IRD) from the high northern latitudes is also shown^{6,7}. EECO, Early Eocene Climatic Optimum. **b**, CaCO_3 mass accumulation rate for the equatorial Pacific^{9,29}. The inset shows the classic CCD record for the equatorial Pacific 50 Myr ago to the present, with the pink-shaded zone

indicating the interval of transient CCD events⁹. Epochs are indicated (abbreviated) to the left of **a** and in the inset. **c**, Benthic $\delta^{18}\text{O}$ record from ODP Site 1218 (the dark green data points are from ref. 8 and yellow ones are from ref. 16). The shaded grey zones are: I, pre-onset of CAE-3; II, onset-to-peak CaCO_3 mass accumulation rate during CAE-3; III, post-peak CaCO_3 mass accumulation rate during CAE-3. Red arrows denote data points defining⁸ a 1.2‰ amplitude of $\delta^{18}\text{O}$ increase ($\Delta\delta^{18}\text{O}$) and thus the hypothesized extreme bipolar glaciation. **d**, **e**, New benthic and planktic $\delta^{18}\text{O}$ records from ODP Site 1260. Median values (white lines) and 1σ variance (shaded envelopes) are shown in **c**, **d** and **e**. Vertical dashed red lines in **d** denote our smaller estimate of $\Delta\delta^{18}\text{O}$ ($\sim 0.6\text{‰ VPDB}$), derived from the maximum and minimum values in the 1σ envelope of the benthic record. Magnetochrons are indicated to the left of **c**.



$\delta^{18}\text{O}_{\text{ice}}$ akin to the average for high altitude glaciers in the temperate zone today ($\geq -15\text{‰}$ SMOW) does the calculated ice volume exceed the modern budget for Antarctica (Fig. 2).

The main reason why the ice budgets of ref. 8 are so large is because their estimate of $\Delta\delta^{18}\text{O}_{\text{benthic}}$ was based on outlying data points and their isotope record is particularly sparse towards the top of magnetochron C19r (note the coring gap on the run up to CAE-3 and inferred glaciation, at the zone I/II boundary in Fig. 1c). To assess this interval in more detail we show an expanded version of our

Figure 2 | Measured increases in $\delta^{18}\text{O}$ in benthic foraminifera ($\Delta\delta^{18}\text{O}_{\text{benthic}}$) and resulting global continental ice volumes for a range of values for the mean $\delta^{18}\text{O}$ of stored ice ($\delta^{18}\text{O}_{\text{ice}}$). Horizontal dashed lines indicate different values for $\Delta\delta^{18}\text{O}_{\text{benthic}}$ (described in the text) and assume $\Delta\delta^{18}\text{O}_{\text{benthic}} = 100\text{‰}$ ice volume (a, equatorial Pacific⁸; b, equatorial Atlantic, this study). Thick vertical lines correspond to different values of $\delta^{18}\text{O}_{\text{ice}}$ (estimated $\delta^{18}\text{O}_{\text{ice}}$ today at -50‰ and for the Middle Eocene at -30‰). Intersections yield resulting calculated ice volume shown by 'isovols' (values shown are $\times 10^6 \text{ km}^3$). The horizontal dashed line labelled b' corresponds to $\Delta\delta^{18}\text{O}_{\text{benthic}} = 0.6\text{‰}$, with 50% of that signal attributable to ice. Present Antarctic ice volume (PAIV)²⁴, Antarctic ice volume at the Last Glacial Maximum (ALGM)³⁰ and total global ice volume at the Last Glacial Maximum (GLGM) shown for reference. SMOW, standard mean ocean water.

foraminiferal $\delta^{18}\text{O}$ series together with accompanying $\delta^{13}\text{C}$, lithological and bulk sediment records in Fig. 3. Excursions to lower values of $\delta^{18}\text{O}$ (by $\sim 0.4\text{‰}$) and $\delta^{13}\text{C}$ (by $1.2\text{--}1.5\text{‰}$) occur in both benthic and planktic records. These are accompanied by prominent decreases in bulk $\delta^{18}\text{O}$ (by 1.8‰) and $\delta^{13}\text{C}$ (by 1.1‰) within magnetochron C19r, across a 30 cm interval containing a highly dissolved foraminiferal assemblage and a well-developed red clay horizon where carbonate content falls sharply from ~ 75 weight percent CaCO_3 to a minimum of ~ 35 weight percent CaCO_3 (Fig. 3f). This event goes undetected in the equatorial Pacific record, presumably because it falls in the core gap at ODP Site 1218 (Fig. 1c) and therefore cannot explain the discrepancy between the two $\Delta\delta^{18}\text{O}_{\text{benthic}}$ signals shown in Fig. 1. On the other hand, our records show that, while ODP Site 1260 remained above the CCD for the vast majority of the study interval, shortly before the CCD deepened in the equatorial Pacific (CAE-3 in Fig. 1), it shoaled in the equatorial Atlantic sufficiently to be readily detectable at a water depth of about 3.0 to 2.5 km. In this respect, our records resemble those for the Eocene/Oligocene boundary where the CCD shoals sufficiently to completely eliminate CaCO_3 burial at ODP

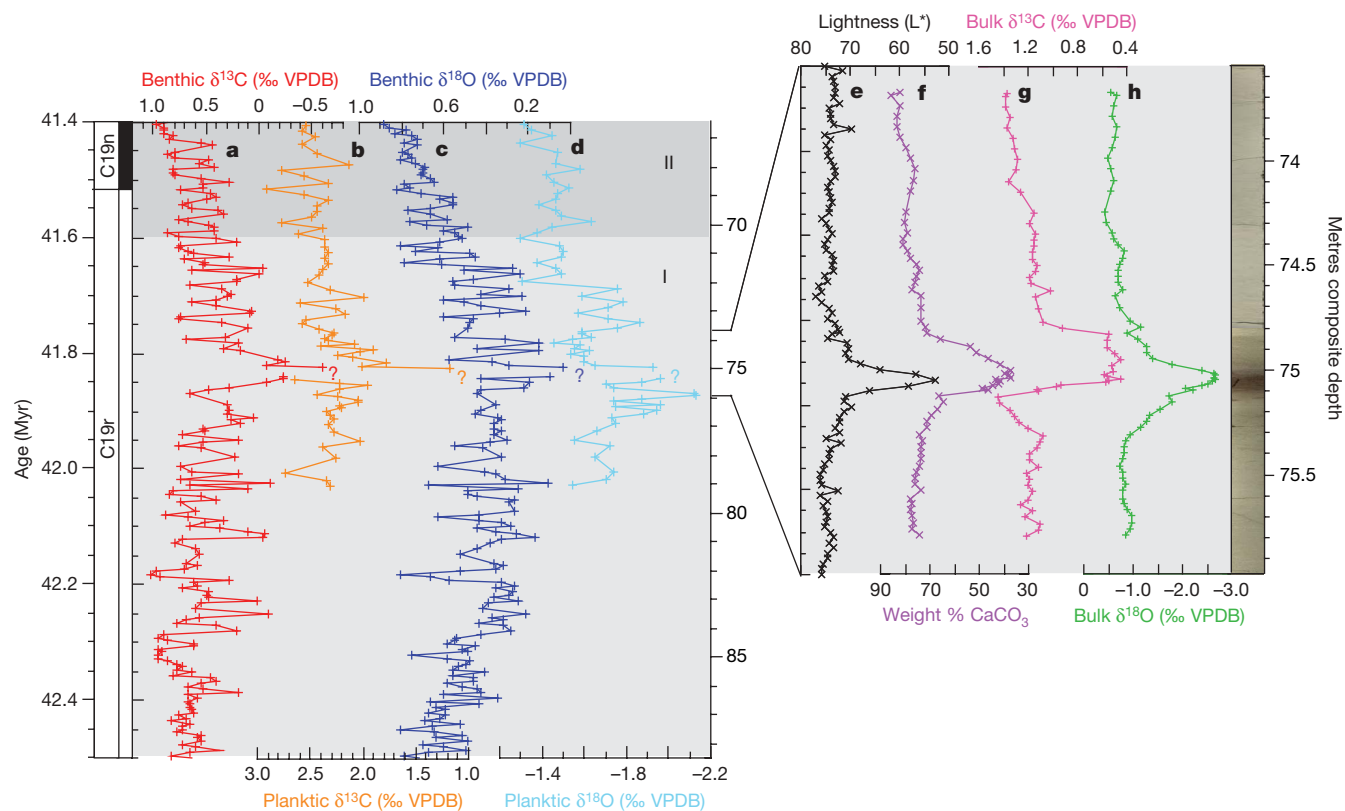


Figure 3 | Palaeoceanographic records from ODP Site 1260 showing the run-up to CAE-3. We note the short-lived interval of CCD shoaling and ocean warming around 75 metres composite depth. **a**, Benthic $\delta^{13}\text{C}$ data. **b**, Planktic $\delta^{13}\text{C}$ data. **c**, Benthic $\delta^{18}\text{O}$ data. **d**, Planktic $\delta^{18}\text{O}$ data. **e**, Bulk

sediment lightness²². **f**, Bulk sediment weight % CaCO_3 . **g**, Bulk sediment $\delta^{13}\text{C}$. **h**, Bulk sediment $\delta^{18}\text{O}$. Question marks in **a–d** denote absence of two to three samples because of the lack of suitable foraminifera for analysis in this interval. Grey shaded zones I and II as in Fig. 1.

Site 1218 for ~200 kyr immediately before the rapid two-step deepening associated with Antarctic glaciation (see figure 1 of ref. 4).

We cannot yet determine whether the C19r event is global in occurrence because of the coring gap at ODP Site 1218 and radiolarian clays and low sedimentation rates ($\leq 0.2 \text{ cm kyr}^{-1}$) at other key drill sites in the Equatorial Pacific¹⁷ and on Shatsky Rise²⁷. But it is distinct from the much longer-lived Middle Eocene Climate Optimum already documented at Site 1260¹⁸ and it is more similar in expression to the well-known intervals of rapid carbon-cycle-led global warming and CCD shoaling in the early Palaeogene. The magnitude of the palaeoceanographic signals that we have documented for magnetochron C19r at Site 1260 is smaller than seen at the Palaeocene/Eocene boundary^{3,10} but at least as large as those marking the Eocene Thermal Maximum 2 (~53 Myr ago)¹¹.

Our findings appear to highlight the instability of Eocene climate and to emphasize close coupling to the carbon cycle as expressed in oceanic carbonate saturation and $\delta^{13}\text{C}$. If the interval of CCD shoaling and isotopic shifts that we document is substantiated elsewhere, our results suggest that carbon-cycle-led instabilities akin to those documented in the early Eocene were not restricted to the Cenozoic climate optimum but also occurred millions of years later when the Earth's baseline climate had cooled substantially and was poised much closer to the threshold of continental glaciation (Fig. 1a). Climate modelling experiments suggest^{12,13} that this threshold would have been crossed earlier in the Southern Hemisphere than in the Northern Hemisphere (Fig. 1a) because of the fundamentally different land-ocean distributions at the two poles. Our work helps to re-align the geologic record with this view.

METHODS

Chronology. We use the published²⁸ palaeomagnetic reversal stratigraphy for ODP Site 1260, which we supplemented by analysing nearly 100 additional samples collected at approximately 20 to 30 cm intervals across each of the magnetic reversals in the studied stratigraphic interval (magnetochrons C20r through C18r). Methods are as in ref. 28.

CaCO₃ and stable isotope data. All data were generated at the National Oceanography Centre, Southampton, UK. Bulk sediment weight %CaCO₃ was measured on small (20–30 mg) discrete samples (1–5 cm spacing), by high-precision coulometry. Our foraminiferal stable isotope ($\delta^{18}\text{O}$ and $\delta^{13}\text{C}$) data were generated using species separates of the benthic and planktic foraminifera *Cibicidoides eoceanus* and *Morozovella lehneri* respectively. Foraminifera were cleaned by ultrasonication before analysis. Benthics were picked from the size range 300–450 μm while planktics were picked from the 250–300 μm size fraction. Bulk stable isotope measurements were generated on splits of the same samples used for CaCO₃ data. All stable isotope measurements were determined using a Europa GEO 20-20 mass spectrometer equipped with an automatic carbonate preparation system. Results are reported relative to the Vienna Pee Dee Belemnite (VPDB) standard with an external analytical precision, based on replicate analysis of an in-house standard calibrated to NBS-19, better than 0.1‰ for $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$. Benthic $\delta^{18}\text{O}$ values that we report from ODP Site 1260 have been adjusted to equilibrium by adding 0.28‰ VPDB, the Palaeogene correction factor for *Cibicidoides*. We plot *Cibicidoides* data from ODP Site 1218 in the same way to aid comparison (Site 1218 data from *Nuttallides truempyi* are reported in the same way as in ref. 8).

Received 16 January; accepted 28 June 2007.

- Miller, K. G., Wright, J. D. & Fairbanks, R. G. Unlocking the ice house: Oligocene-Miocene oxygen isotopes, eustasy, and margin erosion. *J. Geophys. Res.* **96**, 6829–6848 (1991).
- Lear, C. H., Elderfield, H. & Wilson, P. A. Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite. *Science* **287**, 269–272 (2000).
- Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
- Coxall, H. K., Wilson, P. A., Pälike, H., Lear, C. H. & Backman, J. Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature* **433**, 53–57 (2005).
- Holbourn, A., Kuhnt, W., Schulz, M. & Erlenkeuser, H. Impacts of orbital forcing and atmospheric carbon dioxide on Miocene ice-sheet expansion. *Nature* **438**, 483–487 (2005).
- Moran, K. et al. The Cenozoic palaeoenvironment of the Arctic Ocean. *Nature* **441**, 601–605 (2006).
- Eldrett, J. S., Harding, I. C., Wilson, P. A., Butler, E. & Roberts, A. P. Continental ice in Greenland during the Eocene and Oligocene. *Nature* **446**, 176–179 (2007).
- Tripathi, A., Backman, J., Elderfield, H. & Ferretti, P. Eocene bipolar glaciation associated with global carbon cycle changes. *Nature* **436**, 341–346 (2005).
- Lyle, M., Olivarez Lyle, A., Backman, J. & Tripathi, A. Biogenic sedimentation in the Eocene equatorial Pacific—The stuttering greenhouse and Eocene carbonate compensation depth. *Proc. ODP Sci. Res.* (eds Wilson, P. A., Lyle, M. & Firth, J. V.) **199**, 1–35 (2005).
- Zachos, J. C. et al. Rapid acidification of the ocean during the Paleocene-Eocene thermal maximum. *Science* **308**, 1611–1615 (2005).
- Lourens, L. J. et al. Astronomical pacing of late Palaeocene to early Eocene global warming events. *Nature* **435**, 1083–1087 (2005).
- DeConto, R. M. & Pollard, D. Northern Hemisphere glaciation in the early Cenozoic? *AGU Fall Meet. abstr.* PP21B-1678 (2006).
- DeConto, R. M. & Pollard, D. Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO₂. *Nature* **421**, 245–249 (2003).
- Zachos, J. C., Breza, J. R. & Wise, S. W. Early Oligocene ice-sheet expansion on Antarctica: Stable isotope and sedimentological evidence from Kerguelen Plateau, southern Indian Ocean. *Geology* **20**, 569–573 (1992).
- Pekar, S. F., Christie-Blick, N., Kominz, M. A. & Miller, K. G. Calibration between eustatic estimates from backstripping and oxygen isotopic records for the Oligocene. *Geology* **30**, 903–906 (2002).
- Lear, C. H., Rosenthal, Y., Coxall, H. K. & Wilson, P. A. Late Eocene to early Miocene ice sheet dynamics and the global carbon cycle. *Paleoceanography* **19**, doi:10.1029/2004PA001039 (2004).
- Shipboard. Scientific Party 2002. Leg 199 Summary. *Proc. ODP Init. Rep.* (eds Lyle, M., Wilson, P. A., Janecek, T. R.) **199**, 1–87 (2002).
- Sexton, P. F., Wilson, P. A. & Norris, R. D. Testing the Cenozoic multi-site composite $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ curves: new Eocene monospecific records from a single locality, Demerara Rise (ODP Leg 207). *Paleoceanography* **21**, PA2019, doi:10.1029/2005PA001253 (2006).
- Pagani, M., Zachos, J. C., Freeman, K. H., Tipler, B. & Bohaty, S. Marked decline in atmospheric carbon dioxide concentration during the Paleogene. *Science* **309**, 600–603 (2005).
- Wolfe, J. A. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 357–375 (Geophys. Monogr. 32, American Geophysical Union, Washington DC, 1985).
- Pekar, S. F., Hucks, A., Fuller, M. & Li, S. Glacioeustatic changes in the Early and Middle Eocene (51–42 Ma): shallow-water stratigraphy from ODP Leg 189 Site 1171 (South Tasman Rise) and deep-sea $\delta^{18}\text{O}$ records. *Geol. Soc. Am. Bull.* **117**, 1081–1093 (2005).
- Shipboard. Scientific Party 2004. Leg 207 Summary. *Proc. ODP Init. Rep.* (eds Erbacher, J., Mosher, D. C. & Malone, M. J.) **207**, 1–89 (2004).
- Pekar, S. F., DeConto, R. M. & Harwood, D. M. Resolving a late Oligocene conundrum: deep-sea warming and Antarctic glaciation. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **231**, 29–40 (2006).
- Lythe, M. B., Vaughan, D. G. & the BEDMAP consortium. BEDMAP: a new ice thickness and subglacial topographic model of Antarctica. *J. Geophys. Res.* **106**, 335–351 (2001).
- Poulsen, C. J., Pollard, D. & White, T. S. General circulation model simulation of the $\delta^{18}\text{O}$ content of continental precipitation in the middle Cretaceous: A model-proxy comparison. *Geology* **35**, 199–202 (2007).
- Schrag, D. P. et al. The oxygen isotopic composition of seawater during the Last Glacial Maximum. *Quat. Sci. Rev.* **21**, 331–342 (2002).
- Shipboard. Scientific Party. Leg 198 Summary. *Proc. ODP Init. Rep.* (eds Bralower, T. J., Premoli-Silva, I. & Malone, M. J.) **198**, 1–148 (2002).
- Suganuma, Y. & Ogg, J. G. Campanian through Eocene Magnetostratigraphy of Sites 1257–1261, ODP Leg 207, Demerara Rise (Western Equatorial Atlantic). *Proc. ODP Sci. Res.* (eds Mosher, D. C., Erbacher, J. & Malone, M. J.) **207**, 1–48 (2006).
- Pälike, H. et al. The heartbeat of the Oligocene climate system. *Science* **314**, 1894–1898 (2006).
- Huybrechts, P. Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. *Quat. Sci. Rev.* **21**, 203–231 (2002).

Acknowledgements This work used samples provided by the Ocean Drilling Program (ODP). The ODP (now IODP) is sponsored by the US National Science Foundation and participating countries under management of the Joint Oceanographic Institutions (JOI), Inc. We thank the shipboard party of ODP Leg 207, M. Bolshaw, M. Cooper and D. Spanner for help with laboratory work, H. Pälike for making data available; S. Gibbs, K. Moriya, H. Pälike, A. Roberts and T. Tyrrell for discussions and comments on an earlier draft and R. Zeebe for his constructive review. Financial support was provided by NERC in the form of a UK IODP grant to P.A.W. and P.F.S. and by NERC and Perkin Elmer in the form of a CASE studentship to K.M.E.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.A.W. (paw1@noc.soton.ac.uk).

LETTERS

Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years

Kenji Kawamura^{1,2,†}, Frédéric Parrenin³, Lorraine Lisiecki⁴, Ryu Uemura⁵, Françoise Vimeux^{6,7}, Jeffrey P. Severinghaus², Manuel A. Hutterli⁸, Takakiyo Nakazawa¹, Shuji Aoki¹, Jean Jouzel⁷, Maureen E. Raymo⁴, Koji Matsumoto^{1,†}, Hisakazu Nakata^{1,†}, Hideaki Motoyama⁵, Shuji Fujita⁵, Kumiko Goto-Azuma⁵, Yoshiyuki Fujii⁵ & Okitsugu Watanabe⁵

The Milankovitch theory of climate change proposes that glacial–interglacial cycles are driven by changes in summer insolation at high northern latitudes¹. The timing of climate change in the Southern Hemisphere at glacial–interglacial transitions (which are known as terminations) relative to variations in summer insolation in the Northern Hemisphere is an important test of this hypothesis. So far, it has only been possible to apply this test to the most recent termination^{2,3}, because the dating uncertainty associated with older terminations is too large to allow phase relationships to be determined. Here we present a new chronology of Antarctic climate change over the past 360,000 years that is based on the ratio of oxygen to nitrogen molecules in air trapped in the Dome Fuji and Vostok ice cores^{4,5}. This ratio is a proxy for local summer insolation⁵, and thus allows the chronology to be constructed by orbital tuning without the need to assume a lag between a climate record and an orbital parameter. The accuracy of the chronology allows us to examine the phase relationships between climate records from the ice cores^{6–9} and changes in insolation. Our results indicate that orbital-scale Antarctic climate change lags Northern Hemisphere insolation by a few millennia, and that the increases in Antarctic temperature and atmospheric carbon dioxide concentration during the last four terminations occurred within the rising phase of Northern Hemisphere summer insolation. These results support the Milankovitch theory that Northern Hemisphere summer insolation triggered the last four deglaciations^{3,10,11}.

Antarctic deep ice cores have provided a wealth of information on glacial–interglacial climate change and atmospheric composition^{6–9}. However, uncertainties of current timescales are large. For the Dome Fuji⁷, Vostok¹² and Dome C¹³ cores, timescales based on inverse methods using ice flow models disagree by up to ~5,000 years (~5 kyr), because of uncertainties in modelled accumulation rate and/or ice thinning¹³. Another approach is orbital tuning, in which a global atmospheric signal ($\delta^{18}\text{O}$ of O_2 , or concentration of CH_4) is matched to presumed orbital forcing with an assumed lag^{14,15}. This approach is generally affected by the difference between the ice age and the gas age, which adds error of the order of 1 kyr to the ice chronology¹⁶, in addition to uncertainty in the assumed lag of several kyr (see below). Here we eliminate most of these uncertainties with a new orbital tuning that uses the O_2/N_2 ratio of trapped air, which is a proxy for the local summer insolation at the deposition sites⁵.

Because O_2 molecules are preferentially excluded from freshly formed air bubbles during the snow–ice transition¹⁷, the O_2/N_2 ratio of air trapped in Antarctic ice is depleted by 5–10‰ on average from the atmospheric ratio. The O_2/N_2 in the Vostok⁵ and Dome Fuji⁴ cores (Fig. 1) varies in accordance with local summer insolation with an amplitude of ~10‰, with stronger insolation leading to more O_2/N_2 depletion. Local summer insolation presumably controls the physical properties of near-surface snow, which in turn determines the magnitude of O_2/N_2 fractionation during the bubble close-off⁵. The O_2/N_2 records of Antarctic inland cores are expected to be in phase with summer solstice insolation. Support for this view comes from the present-day observation at Dome Fuji of maxima within several days of the solstice in snow temperature (10-cm depth) and vertical temperature gradient (between 10 and 50 cm) (Supplementary Fig. 1), both of which are important factors for snow metamorphism⁵. Positive feedback between snow grain size and absorption of solar radiation, as well as a negative relationship between surface albedo and solar zenith angle, increases the importance of the solstice⁵.

Because of the physical link, the O_2/N_2 orbital tuning is distinct from traditional orbital tuning for ice and marine cores using global signals. The Dome Fuji and Vostok⁵ O_2/N_2 records show no detectable climatic imprint (Supplementary Information), a feature that makes this proxy an ideal tuning tool. In particular, the O_2/N_2 records lack a nonlinear 100-kyr response, which has to be ignored or modelled during traditional orbital tuning. Our timescale is thus ideal for examining terminations. Another advantage is that our method gives the age of ice despite its use of the gas record, because the O_2/N_2 fractionation during bubble formation is determined by the physical properties of surrounding ice.

We derive chronological tie points between the O_2/N_2 record and local summer solstice insolation by peak-to-peak matching (Methods), to modify the original glaciological timescale⁷ (DFGT-2003). We can use O_2/N_2 data only from the depths in which air is occluded as clathrate hydrates⁴, which restricts the youngest tie point to ~82 kyr ago (all dates given are relative to AD 2000). The DFGT-2003 timescale, which has a constraint at ~41 kyr ago, is unchanged for 0–41.5 kyr ago and its agreement with Greenland timescales is probably within 1 kyr, based on a comparison with the Byrd-core isotope record on the GISP2 timescale¹⁸. We linearly interpolated the DFGT-2003 timescale between all the O_2/N_2 tie points, as well as between 41.5 and 82 kyr ago.

¹Center for Atmospheric and Oceanic Studies, Graduate School of Science, Tohoku University, Sendai 980-8578, Japan. ²Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0244, USA. ³Laboratoire de Glaciologie et Géophysique de l'Environnement, CNRS/UJF, 54 rue Molière, 38400 Grenoble, France. ⁴Department of Earth Sciences, Boston University, 685 Commonwealth Avenue, Boston, Massachusetts 02215, USA. ⁵National Institute of Polar Research, Research Organization of Information and Systems, 1-9-10 Kaga, Itabashi-ku, Tokyo 173-8515, Japan. ⁶Institut de Recherche pour le Développement (IRD), UR Great Ice. ⁷IPSL/LSCE, Laboratoire des Sciences du Climat et de l'Environnement, UMR CEA-CNRS-UVSQ, CE Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette, France. ⁸British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK. [†]Present addresses: National Institute of Polar Research, Research Organization of Information and Systems, 1-9-10 Kaga, Itabashi-ku, Tokyo 173-8515, Japan (K.K.); Japan Meteorological Agency, 1-3-4 Otemachi, Chiyoda-ku, Tokyo 100-8122, Japan (K.M.); Japan Atomic Energy Agency, Tokai-mura, Ibaraki 319-1195, Japan (H.N.).

We applied our method also to the Vostok core using published data^{5,12} (Fig. 1). The new Vostok timescale (Vko-2006) resulted in an agreement with the Dome Fuji O_2/N_2 timescale to within 1 kyr for most of the overlapping period (~ 200 – 340 kyr ago), indicating that our method is robust. A large discrepancy was found only where the Dome Fuji O_2/N_2 data are too noisy to reliably constrain the timescale (220–230 kyr ago, Fig. 1b and c). To improve and extend the Dome Fuji O_2/N_2 timescale, we adopted tie points at 221.2, 230.8 and 334.9 kyr ago (Fig. 1c) from the Vostok core, by transferring them to the Dome Fuji core by visual matching of the isotopic records^{7,12} (the final Dome Fuji timescale is named DFO-2006). We note that an independent study¹⁹, limited to Vostok, used new¹⁹ and published⁵ O_2/N_2 data to derive a similar timescale, and discussed palaeoclimate implications.

The dating uncertainties at the tie points (Fig. 1c) due to the noise in the O_2/N_2 data (for example, from non-orbital natural variation, experimental error and gas loss correction) are estimated with a Monte Carlo method to range from 0.8 to 2.9 kyr (Methods). The uncertainty between the tie points originating in the DFGT-2003 timescale and its linear interpolation is estimated using an autocorrelated random walk model (Fig. 1e; Supplementary Information). The accuracy of the DFO-2006 timescale is validated through comparison with four radiometric-dated time markers spanning 92–131 kyr ago, all of which agree within the uncertainty (Supplementary Information).

This accurate chronology permits a critical examination of the phase relationship between Antarctic climate and orbital variations (Fig. 2). We examine the $\delta^{18}O_{ice}$ record of the Dome Fuji core and a preliminary reconstruction of 'site temperature' (ΔT_{site}) based on an isotopic inversion, which accounts for the influence of seawater $\delta^{18}O$ (hereafter $\delta^{18}O_{sw}$) and vapour source temperature²⁰ (Methods). Our chronology also indirectly gives the timing of the CO_2 rise at terminations, which occurs within ~ 1 kyr of the increase in Antarctic temperature²¹.

A striking feature of Dome Fuji $\delta^{18}O_{ice}$ and ΔT_{site} records is that they appear to reflect every precessional-band variation of the northern insolation curve (Fig. 2), even for Marine Isotope Stages (MIS) 6c and 8c where changes are small, as well as the weak variation in MIS 3.

The result confirms earlier observations^{6,22} of the similarity of the patterns of northern insolation and the Antarctic climate, and further provides the phasing between them. The $\delta^{18}O_{ice}$ power spectrum (Fig. 3) shows strong power in the precession band (23-kyr period) with a mean lag of 1.0 ± 0.5 kyr behind precession. The site temperature ΔT_{site} shows smaller obliquity and precession components than $\delta^{18}O_{ice}$, and the precession component is as strong as the obliquity component (Fig. 3). The lag of ΔT_{site} behind precession is 1.8 ± 0.5 kyr. At 41-kyr period, $\delta^{18}O_{ice}$ and ΔT_{site} lag behind obliquity by 2.1 ± 0.7 and 4.7 ± 1.1 kyr, respectively. Importantly, the southern summer insolation (Fig. 2) is in anti-phase (or is completely out of phase) with Antarctic climate, so does not appear to have a strong influence on the orbital-scale Antarctic climatic changes. These phasings suggest that the Antarctic climate on orbital timescales is paced by northern summer insolation presumably through northern ice sheet variation, although minor influence by southern insolation cannot be entirely ruled out. Our data therefore disfavour hypotheses that call for early Southern Hemisphere warming by southern summer insolation to trigger the northern deglaciation²³.

Another remarkable feature of the data is that the last four Antarctic terminations occur within the rising phase of northern summer insolation. We examine the onset of rapid warming towards interglacial warmth, excluding prior gradual warming for terminations II and III (Fig. 4). Because the gradual warming is attributable to a linear climatic response to northern summer insolation for all climatic substages, they should be separated from terminations when examining hypotheses^{10,14,23,24}. The onsets are distinctly defined in the Antarctic climate records for terminations I, II and IV (Fig. 4). We identify the onset of termination III as the start of the continuous warming at 247 kyr ago after a brief temperature reversal, but our conclusion is unchanged if the prior small peak is included. These onsets occur 6, 3, 7 and 2 kyr later than the $65^\circ N$ summer solstice insolation minima, respectively, for terminations I, II, III and IV (Supplementary Table 1), and the entire duration of these warmings also fit within the rising phase of northern summer insolation (Fig. 4). The timing of the last four terminations is thus fully consistent with Milankovitch theory^{3,10}. One possible explanation for the difference in the timings of terminations relative to insolation is the hypothesis

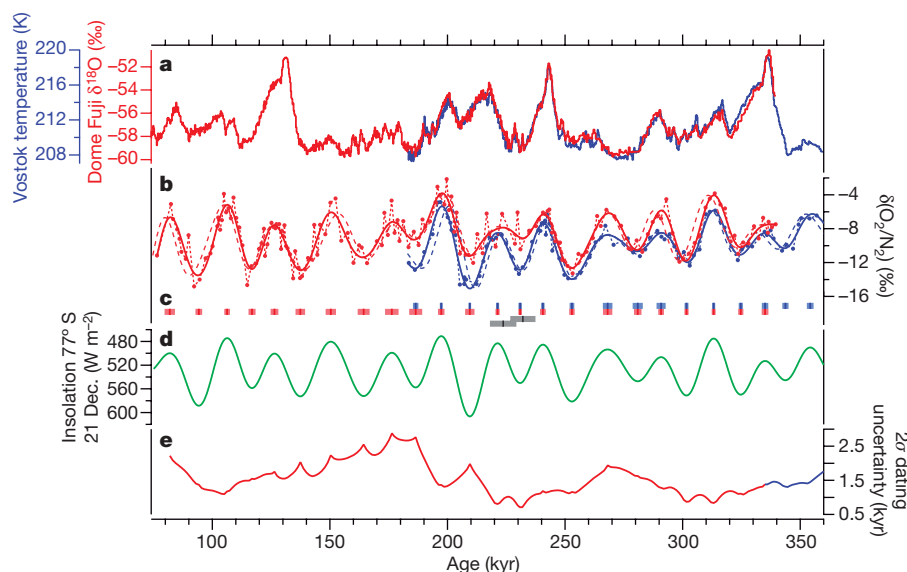


Figure 1 | Orbital tuning of the Dome Fuji and Vostok timescales using O_2/N_2 records. **a**, Dome Fuji $\delta^{18}O_{ice}$ (red, ref. 7) and Vostok temperature¹² (blue, converted from δD) on the respective O_2/N_2 timescales (DFO-2006 and Vko-2006). **b**, O_2/N_2 records of the Dome Fuji⁴ (filled red circles) and Vostok⁵ (filled blue circles) cores and filtered curves (solid lines) on the O_2/N_2 timescales, and filtered curves on the original glaciological timescales (dashed lines). **c**, Age tie points for the Dome Fuji (red) and Vostok (blue)

cores with 2σ error bars. The tie points at 221.2, 230.8 and 334.9 kyr ago for DFO-2006 are adopted from the Vostok data (see text). Two discarded points of the Dome Fuji core are shown (grey markers). **d**, Summer solstice insolation at $77^\circ S$ as the tuning target (inverted axis scale). **e**, 2σ dating uncertainty of DFO-2006 (red) and Vko-2006 (blue, for oldest part) (see Supplementary Information).

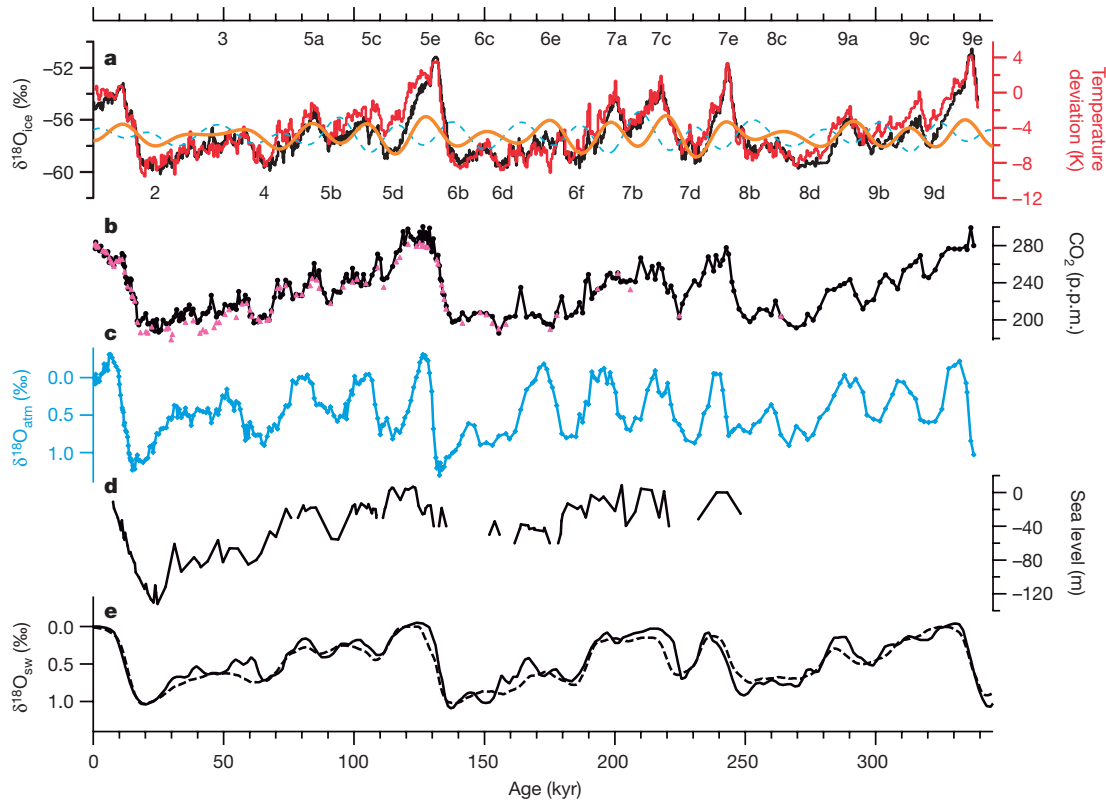


Figure 2 | Comparison of Dome Fuji climatic and atmospheric records with insolation and sea level records. **a**, Dome Fuji $\delta^{18}\text{O}_{\text{ice}}$ (black, ref. 7) and temperature deviation (ΔT_{site}) from the mean of the last 10 kyr (red, this study) on the DFO-2006 timescale, and summer solstice insolation at 65°N (orange) and 65°S (light blue dashed line). The insolation curves are scaled to match 65°N insolation with low-pass filtered $\delta^{18}\text{O}$ (not shown) for the MIS 5b–a transition. Corresponding MIS numbers²⁷ are shown without implying their exact timings. The ΔT_{site} and $\delta^{18}\text{O}_{\text{ice}}$ records are scaled to visually match the changes from the LGM to early Holocene. **b**, Atmospheric CO_2 concentration from the Dome Fuji core by wet extraction⁸ (black circles) and dry extraction (pink triangles, this study). (Two wet-extraction

CO_2 outliers⁸ (at 149.8 and 365.0 m, in Holocene ice) are excluded from this plot. Experimental error rather than chemical reaction in meltwater⁸ may be suspected for these samples, which were measured on the first day of ~ 80 measurement days.) Note that the dry extraction values are too low in the transition zone from air bubbles to clathrate hydrates (for ~ 30 – 55 kyr ago), and the wet-extraction values are too high in the LGM⁸ (around 20 kyr ago). **c**, Dome Fuji $\delta^{18}\text{O}_{\text{atm}}$ record (this study, inverted axis scale). **d**, Sea level reconstruction based on radiometric dating of fossil corals with open system correction²⁸. **e**, Orbitally tuned $\delta^{18}\text{O}_{\text{sw}}$ (inverted axis scale) reconstructed from isotope records of marine sediment cores through regression analyses²⁹ (solid line) and ice sheet modelling³⁰ (dashed line).

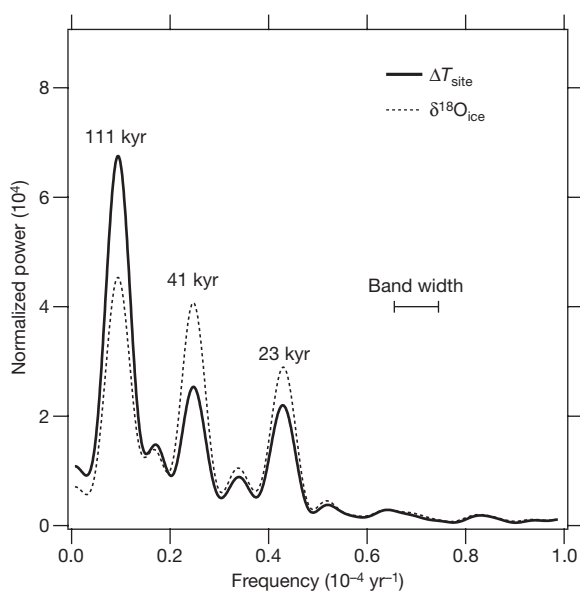


Figure 3 | Power spectra of $\delta^{18}\text{O}$ and ΔT_{site} records calculated by the Blackman–Tukey method with 50% lags. The results (obtained using AnalySeries software) are normalized so that the integrated power over periods longer than 10 kyr becomes unity. Periods of three spectral peaks are shown.

that obliquity and ice volume may modulate the exact phasing between climate and northern summer insolation¹¹.

The onset of the last Antarctic termination was possibly linked to the 19-kyr meltwater pulse, which ended the Last Glacial Maximum (LGM)^{2,25,26}. If the same mechanism holds for the previous terminations, the initial rapid ice volume decrease is predicted to coincide with or precede the onset of Antarctic terminations. A U–Th-dated marine core places the onset of $\delta^{18}\text{O}_{\text{sw}}$ change at ~ 138 kyr ago²³ for termination II, in agreement with the Antarctic timing (137.0 ± 2.2 kyr ago). Therefore, contrary to hypotheses ascribing the trigger of glacial terminations to CO_2 (ref. 14), obliquity²⁴, or southern summer insolation²³, our chronology implicates northern summer insolation as the primary trigger^{3,10,11}.

The major part of post-interglacial Antarctic cooling (MIS 9e–d, 7e–d, and 5e–d transitions) also coincides with decreasing 65°N summer solstice insolation (Fig. 4). Comparison of the Dome Fuji ΔT_{site} with the CO_2 concentration and a radiometric-dated sea level reconstruction (Fig. 2d) as well as orbitally tuned $\delta^{18}\text{O}_{\text{sw}}$ records (Fig. 2f) reveals that the Antarctic cooling for the MIS 5e–d and 7e–d transitions began earlier by several millennia than the corresponding CO_2 and sea level drops (Fig. 2). This suggests that post-interglacial cooling began in the Northern Hemisphere with ice area growth and was transferred to Antarctica quickly through modulation of poleward heat transport and methane concentration decrease⁹—before the reduced CO_2 forcing, or the sea level drop caused by northern ice volume growth, became significant.

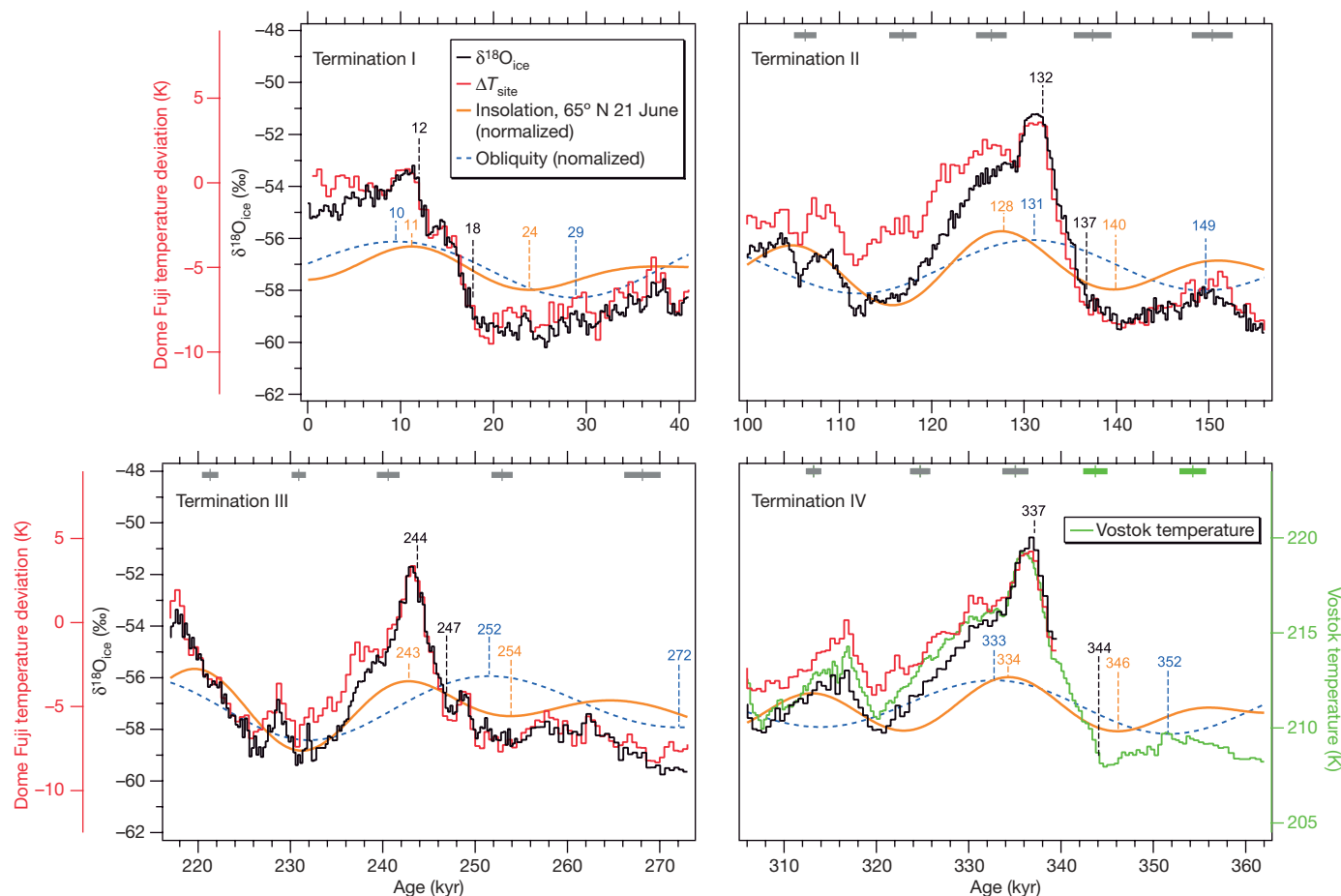


Figure 4 | Comparison of Antarctic parameters with insolation and obliquity around the last four terminations. Shown are Dome Fuji $\delta^{18}\text{O}_{\text{ice}}$ (black, ref. 7) and temperature ΔT_{site} (red, this study) on the DFO-2006 timescale compared with 65°N summer solstice insolation (orange) and obliquity (blue dashed line). The Vostok temperature¹² on the Vko-2006

timescale is also plotted for termination IV (green). Numbers indicate key timings (minima and maxima of insolation and obliquity, and onset and end of $\delta^{18}\text{O}_{\text{ice}}$ increases). The age tie points with 2σ uncertainties are shown at the top.

The phasing between the isotopic composition of atmospheric molecular oxygen ($\delta^{18}\text{O}_{\text{atm}}$) and northern summer insolation varies by ± 6 kyr (Fig. 2), showing that our timescale is inconsistent with $\delta^{18}\text{O}_{\text{atm}}$ -orbital tuning¹⁴. In addition, prior studies have used $\delta^{18}\text{O}_{\text{atm}}$ as a proxy for $\delta^{18}\text{O}_{\text{sw}}$ and thus sea level^{6,14}. However, the fact that $\delta^{18}\text{O}_{\text{atm}}$ lagged the 19-kyr meltwater pulse by 4 kyr indicates a large change of the fractionation between atmospheric and seawater $\delta^{18}\text{O}$ (known as the Dole effect) in the interval 19–15 kyr ago, creating the false impression⁶ that the onset of sea level rise lagged the Antarctic temperature and CO_2 increases. The Dole effect also drove $\delta^{18}\text{O}_{\text{atm}}$ change after ~ 7 kyr ago, when $\delta^{18}\text{O}_{\text{atm}}$ shows a clear shift to glacial-like values despite the stable sea level (Fig. 2). The DFO-2006 timescale reveals another Dole-effect-forced lead of $\delta^{18}\text{O}_{\text{atm}}$ to $\delta^{18}\text{O}_{\text{sw}}$ by ~ 10 kyr at the MIS 5e–d transition (Fig. 2). These observations indicate that the Dole effect involves complex glacial–interglacial changes, and thus orbital tuning and sea level reconstruction based on $\delta^{18}\text{O}_{\text{atm}}$ (ref. 14) are not accurate enough to assess the causality of glacial cycles³. Rather, $\delta^{18}\text{O}_{\text{atm}}$ should be used for reconstruction of the past Dole effect based on the accurate timescale, which is of biogeochemical interest⁵.

In summary, the mean phasing of Antarctic climate, as well as the timing of the last four terminations and three post-interglacial coolings, are consistent with the hypothesis that high northern latitude summer insolation is the trigger of glacial–interglacial cycles. The role of CO_2 as conveyor and amplifier of the orbital input should be quantified with climate models run using our new timescale; this quantification is important for future climate change predictions. Our timescale should be validated further with new radiometric

age markers, as well as by process studies for complete understanding of the physical link between O_2/N_2 and local insolation. With future O_2/N_2 measurements, it may be possible to apply this method to the Dome Fuji and Dome C cores for termination V and older terminations, to investigate the phasing of climate and atmospheric composition with respect to orbital forcing further back in time.

METHODS SUMMARY

All gas measurements on the Dome Fuji core were carried out at Tohoku University. Ice-core air was extracted with a melting technique⁸ and measured with a mass spectrometer (Finnigan MAT Delta-S) with 1σ uncertainty of 0.2‰, 0.02‰ and 0.04‰ for $\delta(\text{O}_2/\text{N}_2)$, $\delta^{15}\text{N}$ of N_2 and $\delta^{18}\text{O}$ of O_2 , respectively. The O_2/N_2 and $\delta^{18}\text{O}$ data were corrected for gravitational enrichment in the firn by subtracting $4 \times \delta^{15}\text{N}$ and $2 \times \delta^{15}\text{N}$, respectively. Because O_2/N_2 in clathrate ice gradually becomes depleted over years in a warm freezer (-25°C) owing to diffusive gas loss⁴, the data were corrected according to the storage period, and outliers were rejected. The new CO_2 data were obtained with a dry extraction method and gravitationally corrected.

Dome Fuji tuning tie points were made for ~ 82 – 335 kyr ago at all maxima and minima of a low-pass filtered O_2/N_2 curve and summer solstice insolation at 77°S . The DFGT-2003 glaciological timescale⁷ was modified according to the tie points (see text). The Vostok tie points were made for ~ 186 – 354 kyr ago to modify the Vko-FGT1 timescale¹². Uncertainties at and between the tie points were estimated by Monte Carlo simulations (Supplementary Information).

Gas age of the Dome Fuji core for CO_2 and $\delta^{18}\text{O}_{\text{atm}}$ was derived by subtracting ice age–gas age difference (Δage) from the DFO-2006 ice age, using a densification model. The Δage depends primarily on accumulation rate, ranging from 1.6 to 4.7 kyr with an accuracy better than 20%.

The Dome Fuji temperature ΔT_{site} was reconstructed with an isotopic inversion using a low-pass filtered deuterium-excess (d-excess) record. Note that the

ΔT_{site} presented here is not intended to resolve millennial changes because of the filtering as well as averaging of ΔT_{site} to 500-yr intervals. Future improvements to the ice core data and traverse snow isotope data from the coast to Dome Fuji will be conducted.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 April; accepted 12 June 2007.

- Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the Earth's orbit: Pacemaker of ice ages. *Science* **194**, 1121–1132 (1976).
- Clark, P. U., McCabe, A. M., Mix, A. C. & Weaver, A. J. Rapid rise of sea level 19,000 years ago and its global implications. *Science* **304**, 1141–1144 (2004).
- Alley, R. B., Brook, E. J. & Anandakrishnan, S. A northern lead in the orbital band: North-south phasing of ice-age events. *Quat. Sci. Rev.* **21**, 431–441 (2002).
- Ikeda-Fukazawa, T. *et al.* Effects of molecular diffusion on trapped gas composition in polar ice cores. *Earth Planet. Sci. Lett.* **229**, 183–192 (2005).
- Bender, M. L. Orbital tuning chronology for the Vostok climate record supported by trapped gas composition. *Earth Planet. Sci. Lett.* **204**, 275–289 (2002).
- Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
- Watanabe, O. *et al.* Homogeneous climate variability across East Antarctica over the past three glacial cycles. *Nature* **422**, 509–512 (2003).
- Kawamura, K. *et al.* Atmospheric CO₂ variations over the last three glacial-interglacial climatic cycles deduced from the Dome Fuji deep ice core, Antarctica using a wet extraction technique. *Tellus B* **55**, 126–137 (2003).
- Spahni, R. *et al.* Atmospheric methane and nitrous oxide of the late Pleistocene from Antarctic ice cores. *Science* **310**, 1317–1321 (2005).
- Raymo, M. E. The timing of major climate terminations. *Paleoceanography* **12**, 577–585 (1997).
- Parrenin, F. & Paillard, D. Amplitude and phase of glacial cycles from a conceptual model. *Earth Planet. Sci. Lett.* **214**, 243–250 (2003).
- Parrenin, F., Remy, F., Ritz, C., Siebert, M. J. & Jouzel, J. New modeling of the Vostok ice flow line and implication for the glaciological chronology of the Vostok ice core. *J. Geophys. Res.* **109**, doi:10.1029/2004JD004561 (2004).
- Parrenin, F. *et al.* The EDC3 agescale for the EPICA Dome C ice core. *Clim. Past Discuss.* **3**, 575–606 (2007).
- Shackleton, N. J. The 100,000-year ice-age cycle identified and found to lag temperature, carbon dioxide, and orbital eccentricity. *Science* **289**, 1897–1902 (2000).
- Ruddiman, W. F. & Raymo, M. E. A methane-based time scale for Vostok ice. *Quat. Sci. Rev.* **22**, 141–155 (2003).
- Bender, M. L. *et al.* Gas age-ice age differences and the chronology of the Vostok ice core, 0–100 ka. *J. Geophys. Res.* **111**, doi:10.1029/2005JD006488 (2006).
- Severinghaus, J. P. & Battle, M. Fractionation of gases in polar ice during bubble close-off: New constraints from firn air Ne, Kr, and Xe observations. *Earth Planet. Sci. Lett.* **244**, 474–500 (2006).
- Blunier, T. & Brook, E. J. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291**, 109–112 (2001).
- Suwa, M. *Chronologies for Ice Cores Constrained by their Gas Records and their Implications for Climate History for the Past 400,000 Years*. Ph.D. thesis. Princeton Univ. (2007).
- Vimeux, F., Cuffey, K. M. & Jouzel, J. New insights into Southern Hemisphere temperature changes from Vostok ice cores using deuterium excess correction. *Earth Planet. Sci. Lett.* **203**, 829–843 (2002).
- Loulergue, L. *et al.* New constraints on the gas age-ice age difference along the EPICA ice cores, 0–50 kyr. *Clim. Past Discuss.* **3**, 435–467 (2007).
- Lorius, C. *et al.* A 150,000-year climatic record from Antarctic ice. *Nature* **316**, 591–596 (1985).
- Henderson, G. M. & Slowey, N. C. Evidence from U-Th dating against Northern Hemisphere forcing of the penultimate deglaciation. *Nature* **404**, 61–66 (2000).
- Huybers, P. & Wunsch, C. Obliquity pacing of the late Pleistocene glacial terminations. *Nature* **434**, 491–494 (2005).
- Yokoyama, Y., Lambeck, K., De Deckker, P., Johnston, P. & Fifield, L. K. Timing of the Last Glacial Maximum from observed sea-level minima. *Nature* **406**, 713–716 (2000).
- Keeling, R. F. & Visbeck, M. Northern ice discharges and Antarctic warming: Could ocean eddies provide the link? *Quat. Sci. Rev.* **24**, 1809–1820 (2005).
- Tzedakis, P. C., Roucoux, K. H., De Abreu, L. & Shackleton, N. J. The duration of forest stages in southern Europe and interglacial climate variability. *Science* **306**, 2231–2235 (2004).
- Thompson, W. G. & Goldstein, S. L. Open-system coral ages reveal persistent suborbital sea-level cycles. *Science* **308**, 401–404 (2005).
- Waelbroeck, C. *et al.* Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21**, 295–305 (2002).
- Bintanja, R., van de Wal, R. S. W. & Oerlemans, J. Modelled atmospheric temperatures and global sea levels over the past million years. *Nature* **437**, 125–128 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Dome Fuji field members for careful drilling and handling of the core, and M. Bender, M. Suwa, R. Alley, P. Huybers, A. Abe-Ouchi, M. Yoshimori, N. Azuma, R. Keeling, Y. Yokoyama, P. Clark, J. Flückiger and W. Ruddiman for discussion and comments. We acknowledge support by a Grant-in-Aid for Creative Scientific Research (to T.N.) and a Grant-in-Aid for Young Scientists (to K.K.) from the Ministry of Education, Science, Sports and Culture, Japan. The Gary Comer Abrupt Climate Change Fellowship and J.P.S. partially supported K.K. during data analysis and writing. M.E.R. acknowledges the support of the US NSF.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.K. (kawamura@nipr.ac.jp).

METHODS

Data correction and selection. Because O_2/N_2 in the clathrate ice becomes gradually depleted over years during sample storage in a warm freezer (for example, -25°C) due to diffusive gas loss⁴, the O_2/N_2 data were corrected according to the storage period assuming a linear O_2/N_2 depletion with time up to 639 days (Supplementary Fig. 2). The samples stored for more than 800 days do not appear to deplete further, thus those data were corrected by the same correction for 639 days. When plotted against age, the corrected O_2/N_2 record shows a smoother and more regular variation than the uncorrected data (Supplementary Fig. 3), and it resembles the independent Vostok record⁵, indicating that the correction is reasonable. We note that using different fits to the depletion trend (exponential function or linear trend throughout all storage durations) result in identical timescales within 2σ uncertainty. Because the post-coring fractionation of bubble ice is not currently understood⁴, the data from shallower depths ($<1,200$ m or younger than ~ 75 kyr ago) are not suitable for dating³¹.

After the correction, outliers in the O_2/N_2 data were rejected on the basis of deviation from a low-pass filtered curve (Supplementary Fig. 4). 28 points out of the original 185 points were rejected with a criterion of 3.2% . The high noise levels for the period younger than ~ 230 kyr ago are attributable to errors in the large gas-loss corrections for the samples stored at -25°C for a long time. Most of the outliers indeed consist of the longest-stored samples. The Dome Fuji O_2/N_2 data can be improved in the future by measuring ice samples stored in a cold freezer (-50°C ; ref. 4) or at Dome Fuji. Measuring O_2/N_2 within a few months after drilling would be ideal for future studies.

Orbital tuning of O_2/N_2 record to local summer solstice insolation. The tuning tie points were made between the O_2/N_2 record and summer solstice insolation at 77°S by peak-to-peak matching (Fig. 1). Note that use of average insolation for a few months around the summer solstice as a tuning target produces identical results within 0.2 kyr. For detecting the maxima and minima in the O_2/N_2 data, as well as for the outlier rejection, the O_2/N_2 data on a glaciological timescale⁷ (DFGT-2003) were linearly interpolated at 0.1 -kyr intervals, and then smoothed by a 559-order finite-duration impulse response (FIR) filter with a Kaiser window (sharp cut-off from 16.7 to 10.0 kyr period; pass-band ripple of less than 1% ; attenuation at stop-band of more than 99.9%), which was designed with the Filter Design and Analysis Tool of the MATLAB software. To avoid phase modulation, the filter was applied twice in the opposite direction (filtfilt function in MATLAB). The peak positions in the filtered curve were used for the matching. We checked the filter performance by applying it to the insolation curve, and confirmed that the peak positions are not altered by more than 0.1 kyr. The timings of the peaks in the filtered O_2/N_2 curve were tied to those in the local summer solstice insolation. Mid-transitions were not used because the relationship between O_2/N_2 and insolation may be nonlinear. Also, the use of mid-transition instead of peaks does not improve the precision by our method. The DFGT-2003 timescale was linearly interpolated between the tie points to derive the O_2/N_2 timescale (DFO-2006). Two tie points at ~ 221 and ~ 230 kyr ago are transferred from the Vostok O_2/N_2 data because the Dome Fuji data are too noisy in this part. We also transferred the tie point at ~ 333 kyr ago from the

Vostok data to better constrain the timescale in the oldest part of the current record.

Ice age–gas age difference. The age of gas in the Dome Fuji core for the $\delta^{18}\text{O}_{\text{atm}}$ and CO_2 records was derived by subtracting an estimate of ice age–gas age difference (Δage) from the DFO-2006 ice age. The Δage was estimated by a semi-empirical firn densification model³² after adjustment of model parameters to reproduce the present density profile (similar to ref. 8). The uncertainty of Δage is estimated by comparing Antarctic Isotope Maxima³³ (AIM) in the Dome Fuji $\delta^{18}\text{O}_{\text{ice}}$ with the abrupt increases of CH_4 (ref. 31), assuming that these two signals are synchronous^{18,33}. The estimated 2σ uncertainties are 8 – 18% of Δage at AIM 1, 2, 8, 14, 17, 23 and 24 (the sampling intervals of CH_4 data for other AIM are too low to usefully constrain the uncertainty). We thus conservatively assign a constant fraction of $\pm 20\%$ as the 2σ uncertainty of Δage (for example, 0.9 kyr in the LGM, 0.5 kyr for the Holocene).

Dry extraction method for CO_2 measurements. The dry extraction method for the CO_2 concentration data has been described elsewhere³⁴. Slight modifications were made as follows. First, the extraction apparatus was pre-conditioned by hot humid air to minimize contamination⁸. Second, a helium cycle cooler (Janis Research) was used instead of liquid helium for cryogenic collection of extracted air. Third, shorter sample tubes were used. Forth, the ice vessel was evacuated for 90 min.

Site temperature reconstruction. An isotopic inversion technique^{20,35} was used to correct the raw isotopic record for both vapour source temperature and marine isotopic composition variations. Owing to the high level of noise in the deuterium excess data especially between 20 and 80 kyr ago, we use a filtered record (passband $> 5,000$ yr, stopband $< 2,000$ yr), as we only discuss the large climate variations. We used a recent marine $\delta^{18}\text{O}_{\text{sw}}$ reconstruction³⁰. We estimate the sensitivities of isotopic values³⁶ to site (Dome Fuji) and source temperatures and marine isotopic composition by using a Rayleigh-based model. We tune this model with the Dome Fuji present-day surface data. Unfortunately, we currently lack isotopic data of surface snow from the coast to Dome Fuji to correctly define the air mass trajectories. Our modelling thus does not offer comprehensive constraints, and future improvements are necessary to interpret ΔT_{site} in detail. Nevertheless, the sensitivities we found are close to those obtained for Vostok²⁰ and Dome C³⁵ (see Supplementary Information).

31. Kawamura, K. *Variations of Atmospheric Components over the Past 340,000 Years from Dome Fuji Deep Ice Core, Antarctica*. Ph.D. thesis. Tohoku Univ. (2001).
32. Barnola, J. M., Pimienta, P., Raynaud, D. & Korotkevich, Y. S. CO_2 -climate relationship as deduced from Vostok ice core: A re-examination based on new measurements and on a re-evaluation of the air dating. *Tellus B* **43**, 83–90 (1991).
33. EPICA. One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature* **444**, 195–198 (2006).
34. Nakazawa, T. et al. Measurements of CO_2 and CH_4 concentrations in air in a polar ice core. *J. Glaciol.* **39**, 209–215 (1993).
35. Stenni, B. et al. An oceanic cold reversal during the last deglaciation. *Science* **293**, 2074–2077 (2001).
36. Uemura, R., Yoshida, N., Kurita, N., Nakawo, M. & Watanabe, O. An observation-based method for reconstructing ocean surface changes using a 340,000-year deuterium excess record from the Dome Fuji ice core, Antarctica. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL019954 (2004).

Hadean diamonds in zircon from Jack Hills, Western Australia

Martina Menneken¹, Alexander A. Nemchin², Thorsten Geisler¹, Robert T. Pidgeon² & Simon A. Wilde²

Detrital zircons more than 4 billion years old from the Jack Hills metasedimentary belt, Yilgarn craton, Western Australia, are the oldest identified fragments of the Earth's crust^{1,2} and are unique in preserving information on the earliest evolution of the Earth. Inclusions of quartz, K-feldspar and monazite in the zircons³, in combination with an enrichment of light rare-earth elements^{4,5} and an estimated low zircon crystallization temperature⁶, have previously been used as evidence for early recycling of continental crust, leading to the production of granitic melts in the Hadean era. Here we present the discovery of microdiamond inclusions in Jack Hills zircons with an age range from $3,058 \pm 7$ to $4,252 \pm 7$ million years. These include the oldest known diamonds found in terrestrial rocks, and introduce a new dimension to the debate on the origin of these zircons and the evolution of the early Earth^{6–10}. The spread of ages indicates that either conditions required for diamond formation were repeated several times during early Earth history or that there was significant recycling of ancient diamond. Mineralogical features of the Jack Hills diamonds—such as their occurrence in zircon, their association with graphite and their Raman spectroscopic characteristics—resemble those of diamonds formed during ultrahigh-pressure metamorphism and, unless conditions on the early Earth were unique, imply a relatively thick continental lithosphere and crust–mantle interaction at least 4,250 million years ago.

Research on >3,900-Myr-old zircon grains from the Jack Hills has mainly focused on their chemical and isotopic compositions. In this study, we have investigated the mineral inclusions in 1,000 randomly chosen zircon grains, collected from a sample of the Jack Hills conglomerate from the initial 'discovery' site¹. U–Pb ages of the zircon grains were determined using a sensitive high-resolution ion microprobe (SHRIMP II), following the analytical procedure described in detail elsewhere^{11,12}. Optical examination revealed that about half of the zircon population contains inclusions larger than $\sim 1 \mu\text{m}$. These inclusions were further investigated by confocal micro-Raman spectroscopy. A Raman spectrum of a mineral reflects both its structural and chemical properties and can thus be used as a fingerprint for its identification. We have identified apatite, quartz, xenotime, monazite, rutile, biotite, amphibole, K-feldspar and plagioclase inclusions in the Jack Hills zircon, some of which have also been recognized in previous studies^{3,4}. In addition, we unambiguously identified diamond inclusions in 45 zircon grains, mostly associated with graphite, but in some cases with apatite and quartz.

The Raman spectrum of diamond is characterized by a single first-order band near $1,332 \text{ cm}^{-1}$ with a typical width (full-width at half-maximum, FWHM) of $\sim 1.7 \text{ cm}^{-1}$, reflecting a triply degenerate C–C stretching vibration with F_{2g} symmetry, and sometimes by a second-order feature near $2,600 \text{ cm}^{-1}$ (ref. 13). The Raman spectra of most of the diamond inclusions ('diamond 1') in the Jack Hills zircons

display the characteristic diamond band at a frequency between $1,329$ and $1,332 \text{ cm}^{-1}$ (Fig. 1a). None of the diamonds show a second-order spectral feature near $2,600 \text{ cm}^{-1}$. Some inclusions ('diamond 2'), however, show an intense band between $1,320$ and $1,327 \text{ cm}^{-1}$ (Fig. 1b) that is further characterized by a larger width compared with that of diamond 1 (Fig. 2). Despite the downshift and broadening, the bands are still symmetric (Fig. 1c, d). At the margin of some diamond inclusions, we obtained spectra that show bands of diamond 1 and 2

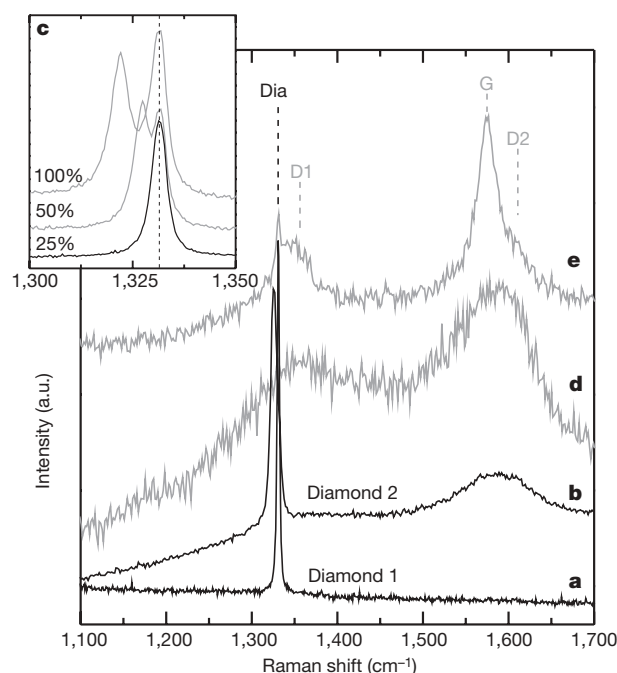


Figure 1 | Representative Raman spectra of diamond and graphite inclusions in zircon from Jack Hills. **a**, Raman spectrum showing the first-order Raman band of diamond near $1,332 \text{ cm}^{-1}$ with a relatively small band width (diamond 1). **b**, Representative spectrum of diamond inclusions that are characterized by a broadened band located between $1,320$ and $1,327 \text{ cm}^{-1}$ (diamond 2). **c**, Raman spectra of a diamond inclusion that is characterized by two bands, representing diamond 1 and 2, measured with different laser intensities. The 100% spectrum was measured with full laser power. Note the blue shift of the left band with decreasing laser power, which demonstrates sample heating effects and indicates that diamond 2 is polycrystalline diamond of variable crystallite sizes. **d**, Raman spectrum of graphite with the broad first-order band (G band) near $1,580 \text{ cm}^{-1}$ and a relative weak disorder band (D1 band) near $1,350 \text{ cm}^{-1}$ (for a 532 nm excitation). **e**, Spectrum of a diamond–graphite composite inclusion. Note that the graphite bands are sharper than those shown in **c** and that an additional disorder band near $1,620 \text{ cm}^{-1}$ is visible (D2 band).

¹Institut für Mineralogie, Westfälische Wilhelms-Universität, Corrensstr. 24, 48149 Münster, Germany. ²Department of Applied Geology, Western Australian School of Mines, Curtin University of Technology, Bentley, Western Australia 6102, Australia.

overlapping each other. In contrast to the characteristic band of diamond 1, the band of diamond 2 shifts towards higher wave-numbers, when decreasing the laser power (Fig. 1c), indicating heating effects induced in diamond 2 by the incident laser. It has been shown¹⁴ that heating effects in diamond powders of grain sizes down to 0.1 μm can cause a size-dependent frequency shift towards 1,320 cm^{-1} that is associated with symmetric band broadening, as a result of different thermal properties of microcrystalline diamond. Hence, the band width–frequency relationship of diamond 2, seen in Fig. 2, reflects the occurrence of polycrystalline diamond of variable grain size that is associated with larger diamonds in some inclusions (Fig. 1c).

A major Raman-active vibration near 1,326 cm^{-1} and an additional band near 1,175 cm^{-1} (ref. 15), characteristic of lonsdaleite (a hexagonal polytype of diamond, which has been reported from impact structures¹⁶), were not observed in the present study (Fig. 1). However, we have identified graphite associated with most diamond inclusions by its broad first-order Raman band near 1,580 cm^{-1} ('G band') and in some cases by its disorder band ('D1 band') near 1,350 cm^{-1} (for a 532 nm excitation) and its second disorder band ('D2 band') near 1,620 cm^{-1} (Fig. 1d and e)¹⁷.

The diamond inclusions have rounded to hexagonal, oval (Fig. 3), angular, or even needle-like shapes. Commonly, graphite is found surrounding diamond, suggesting that it formed after the diamond as a result of retrograde transformation. Diamonds that are exposed at the polished surfaces of the grains can be recognized by their bright cathodoluminescence (CL) intensity, distinguishing them from exposed graphite which gives no CL signal (Figs 3a and d). The secondary electron image in Fig. 3e shows a relatively large, exposed diamond aggregate (giving the spectrum shown in Fig. 1a), which

was originally surrounded by graphite that has now been polished away.

The presence of diamond inclusions in zircon raises the question of whether these diamonds could be the result of contamination from the De Beers diamond polishing powder used in preparing the zircon mounts. However, this can be ruled out as (1) some diamond inclusions are completely enclosed by zircon (Fig. 3b and c), (2) most diamonds reveal variable spectral features that clearly distinguish them from those of the diamond particles in the polishing powder (Fig. 2), (3) the size of the diamond inclusions ranges between 6 and 70 μm so that they are mostly larger than the diamond particles in the polishing materials (<12 μm), and (4) almost all of the diamonds are associated with graphite (Fig. 1d), which is not present in the polishing powder.

The diamond-bearing zircon crystals are characterized by highly variable internal structures in CL images, typical of the whole Jack Hills population^{5,18}. These include simple oscillatory growth zones (Fig. 3d) that are typical of zircon crystals grown from a melt or fluid phase, and irregularly curved growth zones (Fig. 3a) that indicate zircon re-equilibration in an aqueous fluid or melt¹⁹. Those zones containing diamond show a distribution of $^{207}\text{Pb}/^{206}\text{Pb}$ ages similar to the average statistical age distribution of the Jack Hills zircon population^{5,20} with a pronounced peak between 3,200 and 3,400 Myr and with $\sim 10\%$ of the population being older than 3,900 Myr (Supplementary Table 1). We found no correlation between the presence of diamond and the age, the internal growth structure, or the Th/U of the host zircon crystals (Supplementary Table 1). Moreover, the polycrystalline microdiamonds (diamond 2) occur in all age groups (Fig. 2). The three oldest diamond-bearing zircon grains gave concordant ages (Myr) of $4,096 \pm 4$ (1 σ), $4,132 \pm 5$ (1 σ) and $4,252 \pm 7$ (1 σ) (Fig. 3d), making them the oldest known diamonds found in terrestrial rocks. The previously known oldest terrestrial diamonds, with a maximum age defined by garnet inclusions at $3,300 \pm 200$ Myr, are diamonds found in 90-Myr-old kimberlites from Finsch and Kimberley in South Africa²¹.

To evaluate the mode of formation of the Jack Hills diamonds, it is useful to compare the observed mineralogical features with those from diamonds formed under known geological conditions. Natural diamonds form under pressures of >3.2 GPa and temperatures between ~ 650 and 1,200 $^{\circ}\text{C}$ in the Earth's interior at a depth of >100 km or during meteoritic impact events²². Until now, diamonds have been found in (1) primitive carbonaceous chondrites, (2) meteoritic impact structures, (3) kimberlites and lamproites, which delivered them from the subcontinental lithosphere, (4) komatiites, and (5) in crustal rocks, typically sedimentary, that have been subjected to ultrahigh-pressure (UHP) metamorphic conditions²².

UHP metamorphic terranes are at present the only known setting where microdiamonds, about 3–50 μm in size, have been reported as inclusions in zircon^{23–25}. Also, secondary graphite is commonly associated with diamonds that formed in UHP terranes^{26,27} owing to the relatively slow exhumation rates compared to the rate of kimberlite magma ascent, which allows the transformation of diamond to graphite. We further note that the Raman spectral features of diamond 1 agree well with those reported for UHP diamonds^{23,25–28}, and are readily distinguished from those of shock-related diamonds and from diamonds found in kimberlites and lamproites (Fig. 2). The latter diamonds are also characterized by a second-order Raman band near 2,600 cm^{-1} that has not been observed in the present study.

So far we have not detected high-pressure minerals other than diamond in the Jack Hills zircon crystals. There is also no evidence that the observed quartz inclusions were derived from its high-pressure polymorph coesite. The overall mineral inclusion assemblage is essentially granodioritic to granitic (Supplementary Table 1), which resembles the mineral paragenesis commonly associated with carbonados. Carbonados are large polycrystalline nodules that consist of euhedral diamond crystals (typically up to 200 μm)

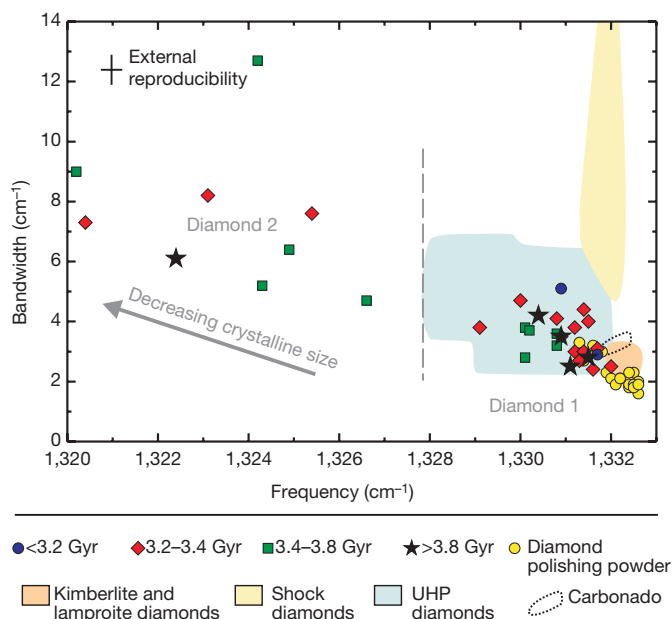


Figure 2 | Raman data from diamond inclusions in the Jack Hill zircons, from diamond polishing particles, and from diamonds of different geological origin. Shown is the width (given as FWHM and corrected for instrumental broadening) of the first-order diamond Raman band as a function of frequency for diamond inclusions in zircon crystals from Jack Hills and for diamond polishing particles. The line broadening and frequency red-shift of diamond 2 reflects sample heating effects as a result of variable grain sizes, as marked by the arrow (see Fig. 1). Note that diamonds with different Raman spectral features (diamond 1 and 2) occur within all age groups of the detrital Jack Hills zircon population. Age groups are defined using the corresponding $^{207}\text{Pb}/^{206}\text{Pb}$ zircon ages (see Supplementary Table 1). Published Raman data^{25,26,28,31} and our own measurements (given in Supplementary Table 2) of diamonds of different origin (shown as coloured fields) are included for comparison.

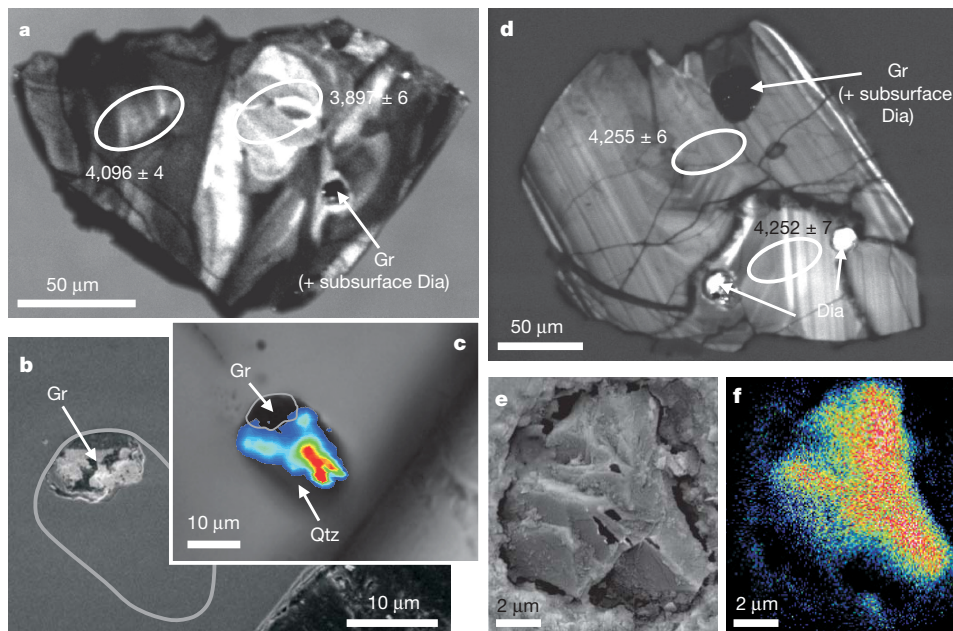


Figure 3 | Mineralogical features of diamond inclusions in Jack Hills zircon samples. a–c, Grain JH15-142; d–f, grain JH3-20. a, Cathodoluminescence (CL) image of grain JH15-142. Gr, graphite. **b,** Backscattered electron (BSE) image of the area around the graphite–diamond composite inclusion shown in **a**. The bright BSE intensity inner region where the inclusion is exposed to the surface is from relict gold coating covering graphite. The dimension of the inclusion, as visible in the transmitted light image shown in **c**, is outlined by the grey curve. **c,** Transmitted light image of the same inclusion as shown in Fig. 1b. The part of the inclusion that is exposed to the surface (see **b**) is outlined by the grey curve. The coloured contours in overlay show the

intensity distribution of the first-order Raman band of diamond near $1,332\text{ cm}^{-1}$. Note that the diamond is included inside the zircon. Qtz, quartz. **d,** CL image of grain JH3-20. Dia, diamond. **e,** Secondary electron image of inclusion 20-3, which is visible in the right corner of the zircon grain shown in **d**. **f,** Carbon distribution of inclusion 20-3 as obtained by energy-dispersive X-ray analysis. Note that carbon signals at the border of the inclusion result from remnant graphite. White ellipses in **a** and **d** mark the SHRIMP analytical sites; numbers show the $^{207}\text{Pb}/^{206}\text{Pb}$ ages (in Myr) with the 1σ uncertainty. Intensity variations in **c** and **f** scale from blue (low intensity) to red (high intensity).

set in a matrix of microcrystalline diamond and are associated with crustal minerals such as quartz, orthoclase, xenotime and zircon²⁹. Their origin is enigmatic, but theories include metamorphism of earliest subducted lithosphere, radioactive transformation of mantle hydrocarbons, and meteoritic impact on a concentrated biomass²⁹. However, the Raman spectral characteristics of carbonados that we have measured are more similar to those of kimberlite- or lamproite-derived diamond rather than to those found in Jack Hills zircon (Fig. 2).

The absence of high-pressure minerals except diamond, together with the occurrence of diamonds in zircons with a range of primary and secondary internal structures (Fig. 3) and the highly variable Th/U ratios (Supplementary Table 1), which do not support formation of zircon at UHP metamorphic conditions, presents a significant problem for the interpretation of the origin of the diamonds and the crystallization history of the zircons. These observations suggest that the diamonds and at least some of their host zircons were formed under different pressure–temperature conditions. It is possible that the diamonds were formed during a single event before or at $4,252 \pm 7$ Myr ago, which is the crystallization age of the oldest diamond-bearing zircon from the Jack Hills conglomerate (Fig. 3d). This hypothesis may be supported by the Hf-isotope compositions previously determined for Jack Hills zircon^{7,30}, which were interpreted to reflect extensive reworking of material separated from the mantle before about 4,140 Myr ago³⁰. Diamonds found in younger zircon grains would then be a result of recycling of the older material. Alternatively, diamonds could have been incorporated in zircon at several periods in early Earth history by repetition of the process that resulted in diamond formation.

Many questions remain regarding the origin of the diamonds and their incorporation in the Jack Hills zircons. However, the comparison of the observed mineralogical features of these diamonds with those formed under known geological conditions, particularly their

Raman spectral characteristics and their occurrence with graphite in zircon, indicates that they share more similarities with UHP metamorphic diamonds than with any other known type. If correct, and unless unknown processes operated during the Earth's earliest history, this would imply the presence of a relatively thick continental lithosphere and that crust–mantle interaction occurred on Earth as early as 4,250 Myr ago.

METHODS SUMMARY

Raman spectroscopy. Laser-Raman spectra were collected with a Jobin Yvon HR800 dispersive Raman spectrometer using the 532 nm line of a 14 mW Nd-YAG laser. The scattered Raman light was analysed with a $100\times$ objective in 180° backscatter geometry and a charged-coupled device (CCD) detector after being dispersed by a grating of 600 grooves mm^{-1} . The diamond inclusions were re-measured for quantitative determination of the position and width of the first-order diamond band with a spectral resolution of 1.1 cm^{-1} (near $1,330\text{ cm}^{-1}$) using a grating of 1,800 grooves mm^{-1} .

Raman images of the diamond-graphite-(quartz) composite inclusion (Fig. 3c) as well as measurements of the diamond powder were performed with an ISA LabRam dispersive Raman spectrometer using the 632.187 nm line of a He-Ne laser and a grating of 1,800 grooves mm^{-1} , yielding a spectral resolution of 1.6 cm^{-1} .

Sensitive high-resolution ion microprobe (SHRIMP). The filtered (O_2^-) beam with intensity between 2 and 3 nA was focused on the surface of samples into $\sim 20\text{ }\mu\text{m}$ spot. Secondary ions were passed to the mass spectrometer operating at a mass resolution ($M/\Delta M$) of $\sim 5,000$. Each analysis was preceded by a 2 min raster to remove the Au coating. The peak-hopping data collection routine consisted of five scans through the mass stations, with signals measured by an ion counting electron multiplier. Pb/U ratios were calibrated using an empirical correlation between Pb^+/U^+ and UO^+/U^+ ratios, normalized to the Curtin University standard (a 564-Myr-old Sri Lankan zircon). The 0.8–1.6% error obtained from the multiple analyses of Pb/U ratio on the standard during individual SHRIMP sessions was added in quadrature to the errors observed in the unknowns.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 April; accepted 6 July 2007.

- Compston, W. & Pidgeon, R. T. Jack Hills, evidence of more very old detrital zircons in Western Australia. *Nature* **321**, 766–769 (1986).
- Wilde, S. A., Valley, J. W., Peck, W. H. & Graham, C. M. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* **409**, 175–178 (2001).
- Maas, R., Kinny, P. D., Williams, I. S., Froude, D. O. & Compston, W. The Earth's oldest known crust: a geochronological and geochemical study of 2900–4200 Ma old zircons from Mt Narryer and Jack Hills, Western Australia. *Geochim. Cosmochim. Acta* **56**, 1281–1300 (1992).
- Peck, W. H., Valley, J. W., Wilde, S. A. & Graham, C. M. Oxygen isotope ratios and rare earth elements in 3.3 to 4.4 Ga zircons: ion microprobe evidence for high $\delta^{18}\text{O}$ continental crust and oceans in the Early Archaean. *Geochim. Cosmochim. Acta* **65**, 4215–4229 (2001).
- Cavosie, A. J., Valley, J. W. & Wilde, S. A. E. I. M. F. Magmatic $\delta^{18}\text{O}$ in 4400–3900 Ma detrital zircons: a record of the alteration and recycling of crust in the Early Archaean. *Earth Planet. Sci. Lett.* **235**, 663–681 (2005).
- Watson, E. B. & Harrison, T. M. Zircon thermometer reveals minimum melting conditions on earliest Earth. *Science* **308**, 841–844 (2005).
- Harrison, T. M. *et al.* Heterogeneous Hadean hafnium: Evidence of continental crust at 4.4 to 4.5 Ga. *Science* **310**, 1947–1950 (2005).
- Valley, J. W., Cavosie, A. J., Fu, B., Peck, W. H. & Wilde, S. A. Comment on “Heterogeneous Hadean hafnium: Evidence of continental crust at 4.4 to 4.5 Ga”. *Science* **312**, 1139a (2006).
- Glikson, A. Comment on “Zircon thermometer reveals minimum melting conditions on earliest Earth” I. *Science* **311**, 779a (2006).
- Nutman, A. P. Comment on “Zircon thermometer reveals minimum melting conditions on earliest Earth” II. *Science* **311**, 779b (2006).
- Compston, W., Williams, I. S. & Meyer, C. U-Pb geochronology of zircons from Lunar Breccia 73217 using a sensitive high mass-resolution ion microprobe. *J. Geophys. Res.* **89**, 525–534 (1984).
- Kennedy, A. K. & de Laeter, J. R. The performance characteristics of the WA SHRIMP II ion microprobe. *US Geol. Surv. Circ.* **1107**, 166 (1994).
- Gillet, P., Hemley, R. J. & McMillan, P. F. Vibrational properties of minerals at high pressures and temperatures. *Rev. Min.* **37**, 525–590 (1998).
- Nachal'naya, T. A. & Andreyev, V. D. Shift of the frequency and Stokes-anti-Stokes ratio of Raman spectra from diamond powders. *Diamond Related Mater.* **3**, 1325–1328 (1994).
- Knight, D. S. & White, W. B. Characterization of diamond films by Raman spectroscopy. *J. Mater. Res.* **4**, 385–393 (1989).
- Frondel, C. & Marvin, U. B. Lonsdaleite, a new hexagonal polymorph of diamond. *Nature* **214**, 587–589 (1967).
- Pasteris, J. D. & Wopenka, B. Raman spectra of graphite as indicators of degree of metamorphism. *Can. Mineral.* **29**, 1–9 (1991).
- Nemchin, A. A., Pidgeon, R. T. & Whitehouse, M. J. Re-evaluation of the origin and evolution of >4.2 Ga zircons from the Jack Hills metasedimentary rocks. *Earth Planet. Sci. Lett.* **244**, 218–233 (2006).
- Geisler, T., Schaltegger, U. & Tomaschek, F. Re-equilibration of zircon in aqueous fluids and melts. *Elements* **3**, 45–51 (2007).
- Pidgeon, R. T. & Nemchin, A. A. High abundance of early Archaean grains and the age distribution of detrital zircons in a sillimanite-bearing quartzite from Mt Narryer, Western Australia. *Precamb. Res.* **150**, 201–220 (2006).
- Richardson, S. H., Gurney, J. J., Erlank, A. J. & Harris, J. W. Origin of diamonds in old enriched mantle. *Nature* **310**, 198–202 (1984).
- Haggerty, S. E. A diamond trilogy: superplumes, supercontinents, and supernovae. *Science* **285**, 851–860 (1999).
- De Corte, K. *et al.* Diamond growth during ultrahigh-pressure metamorphism of the Kokchetav Massif, northern Kazakhstan. *Island Arc* **9**, 428–438 (2000).
- Dobrzhinetskaya, L. F. *et al.* Focused ion beam technique and transmission electron microscope studies of microdiamonds from the Saxonian Erzgebirge, Germany. *Earth Planet. Sci. Lett.* **210**, 399–410 (2003).
- Dobrzhinetskaya, L. F. *et al.* Synchrotron infrared and Raman spectroscopy of microdiamonds from Erzgebirge, Germany. *Earth Planet. Sci. Lett.* **248**, 340–349 (2006).
- Perraki, M., Proyer, A., Mposkos, E., Kaindl, R. & Hoinkes, G. Raman micro-spectroscopy on diamond, graphite and other carbon polymorphs from the ultrahigh-pressure metamorphic Kimi Complex of the Rhodope Metamorphic Province, NE Greece. *Earth Planet. Sci. Lett.* **241**, 672–685 (2006).
- Mposkos, E. & Krohe, A. Pressure-temperature-deformation paths of closely associated ultra-high-pressure (diamond-bearing) crustal and mantle rocks of the Kimi complex: implications for the tectonic history of the Rhodope Mountains, Greece. *Can. J. Earth Sci.* **43**, 1755–1776 (2006).
- Korsakov, A. V., Vandenabeele, P. & Theunissen, K. Discrimination of metamorphic diamond populations by Raman spectroscopy (Kokchetav, Kazakhstan). *Spectrochim. Acta A* **61**, 2378–2385 (2005).
- Heaney, P. J., Vicenzi, E. P. & De, S. Strange diamonds: The mysterious origins of carbonado and framesite. *Elements* **1**, 85–89 (2005).
- Amelin, Y., Lee, D.-C., Halliday, A. N. & Pidgeon, R. T. Nature of the Earth's earliest crust from hafnium isotopes in single detrital zircons. *Nature* **399**, 252–255 (1999).
- El Goresy, A. *et al.* In situ discovery of shock-induced graphite-diamond phase transition in gneisses from the Ries Crater, Germany. *Am. Miner.* **86**, 611–621 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by a Curtin University grant to A.A.N. and S.A.W. We further acknowledge the Deutsche Forschungsgemeinschaft for financial support. We also wish to thank E. Scherer, F. Tomaschek and I. Fitzsimons for discussions on earlier versions of the manuscript and J. Schlüter from the Mineralogical Museum of the University of Hamburg and A. Bischoff for providing diamond samples for comparative Raman measurements.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.M. (m_menn03@uni-muenster.de) or T.G. (tgeisler@nwz.uni-muenster.de).

METHODS

Sample preparation. Zircon grains (1,000) were hand-picked from a previously prepared $>135\ \mu\text{m}$ concentrate and mounted onto double-sided adhesive tape, along with pieces of the Curtin University Sri Lankan gem zircon (CZ3) used as standard for the SHRIMP measurements. They were enclosed in epoxy resin disks, ground so as to effectively cut all zircon grains in half, and polished with $<12\ \mu\text{m}$ De Beers diamond polishing powder.

Raman spectroscopy. Laser-Raman spectra of the inclusions were collected with a Jobin Yvon HR800 dispersive Raman spectrometer using the 532 nm line of a Nd-YAG laser with a laser power of 14 mW. The scattered Raman light was analysed with a CCD detector after being dispersed by a grating of 600 grooves mm^{-1} . A $100\times$ objective with a numerical aperture of 0.9 was used on a BX-40 microscope. The diamond inclusions were re-measured with a grating of 1,800 grooves mm^{-1} for quantitative determination of the position and width of the first-order diamond band near $1,332\ \text{cm}^{-1}$. With the 1,800 grooves mm^{-1} grid, the spectral resolution near $1,332\ \text{cm}^{-1}$ was $1.1\ \text{cm}^{-1}$, as measured by neon emission lines. The following equation was used to correct for the effect of the finite slit width on measured band width (given as FWHM)³²:

$$\Gamma = \Gamma_m [1 - (S/\Gamma_m)^2]$$

where Γ , Γ_m and S are the corrected band width of the first-order diamond band, the measured band width, and the spectral slit width, respectively. The frequency was calibrated using the first-order Si line at $520.7\ \text{cm}^{-1}$ and lines from a neon lamp. The standard errors from the fitting procedure of both the Raman frequency and width were usually less than $0.2\ \text{cm}^{-1}$.

We checked for possible heating effects induced by the incident laser power on the Raman spectral features. Larger diamond inclusions (diamond 1) do not show any heating effects as checked by (1) reducing the laser power with filters and (2) measuring the integral intensity of the Stokes and anti-Stokes component in the diamond Raman spectra, which allows the calculation of the local heating temperature¹⁴ (see Fig. 1c). However, some diamonds (diamond 2) show temperature-induced spectral changes, which are used to discriminate between the two types of diamond (see Fig. 1c and Fig. 2).

Raman images of the diamond-graphite-(quartz) composite inclusions as well as measurements of the diamond powder were performed with an ISA LabRam dispersive Raman spectrometer using the 632.187 nm line of a He-Ne laser and a grating of 1,800 grooves mm^{-1} . Raman images were obtained after the SHRIMP measurements on re-polished samples. The spectral resolution was $1.6\ \text{cm}^{-1}$ near $1,332\ \text{cm}^{-1}$. The lateral spatial resolution was about $1\text{--}2\ \mu\text{m}$. For imaging, a confocal hole of $500\ \mu\text{m}$ was used.

Sensitive high-resolution ion microprobe (SHRIMP). The SHRIMP methodology follows analytical procedure described elsewhere^{11,12}. The filtered (O_2^-) beam with intensity between 2 and 3 nA was focused on the surface of samples into an $\sim 20\ \mu\text{m}$ spot. Secondary ions were passed to the mass spectrometer operating at a mass resolution ($M/\Delta M$) of $\sim 5,000$. Each analysis was preceded by a 2 min raster to remove the Au coating. The peak-hopping data collection routine consisted of five scans through the mass stations, with signals measured by an ion counting electron multiplier. Pb/U ratios were calibrated using an empirical correlation between Pb^+/U^+ and UO^+/U^+ ratios, normalized to the 564-Myr-old Sri Lankan zircon CZ3³³. The 0.8–1.6% error obtained from the multiple analyses of Pb/U ratio on the standard during individual SHRIMP sessions was added in quadrature to the errors observed in the unknowns. The initial data reduction was done using the SQUID add-in for Microsoft Excel³⁴, and Isoplot³⁵ was applied for further age calculations.

32. Tanabe, K. & Hiraishi, J. Correction of finite slit width effects on Raman line widths. *Spectrochim. Acta A* **36**, 341–344 (1980).
33. Pidgeon, R. T., Furlong, D., Kennedy, A. K., Nemchin, A. A. & van Bronswijk, W. Calibration of zircon standards for the Curtin SHRIMP. *US Geol. Surv. Circ.* **1107**, 251 (1994).
34. Ludwig, K. *Users Manual for Squid 1.02* (Special Publication 1a, Berkeley Geochronology Center, Berkeley, California, 2001).
35. Ludwig, K. *Users manual for Isoplot/Ex rev. 2.49* (Special Publication 2, Berkeley Geochronology Center, Berkeley, California, 2001).

A new species of great ape from the late Miocene epoch in Ethiopia

Gen Suwa¹, Reiko T. Kono², Shigehiro Katoh³, Berhane Asfaw⁴ & Yonas Beyene⁵

With the discovery of *Ardipithecus*, *Orrorin* and *Sahelanthropus*^{1–7}, our knowledge of hominid evolution before the emergence of Pliocene species of *Australopithecus*^{8,9} has significantly increased, extending the hominid fossil record back to at least 6 million years (Myr) ago. However, because of the dearth of fossil hominoid remains in sub-Saharan Africa spanning the period 12–7 Myr ago, nothing is known of the actual timing and mode of divergence of the African ape and hominid lineages. Most genomic-based studies suggest a late divergence date—5–6 Myr ago and 6–8 Myr ago for the human–chimp and human–gorilla splits, respectively^{10–14}—and some palaeontological and molecular analyses hypothesize a Eurasian origin of the African ape and hominid clade^{15,16}. We report here the discovery and recognition of a new species of great ape, *Chororapithecus abyssinicus*, from the 10–10.5-Myr-old deposits of the Chorora Formation at the southern margin of the Afar rift. To the best of our knowledge, these are the first fossils of a large-bodied Miocene ape from the African continent north of Kenya. They exhibit a gorilla-sized dentition that combines distinct shearing crests with thick enamel on its ‘functional’ side cusps. Visualization of the enamel–dentine junction by micro-computed tomography reveals shearing crest features that partly resemble the modern gorilla condition. These features represent genetically based structural modifications probably associated with an initial adaptation to a comparatively fibrous diet. The relatively flat cuspal enamel–dentine junction and thick enamel, however, suggest a concurrent adaptation to hard and/or abrasive food items. The combined evidence suggests that *Chororapithecus* may be a basal member of the gorilla clade, and that the latter exhibited some amount of adaptive and phyletic diversity at around 10–11 Myr ago.

The Chorora Formation represents the earliest known record of sedimentation within the Afar rift of Ethiopia¹⁷. Perilacustrine sediments and their palaeontological contents were first reported in the 1970s (ref. 18) and estimated to be between 10 and 11 Myr old^{19,20}. More recent re-evaluations have confirmed this chronology from both radioisotopic dating and biochronological considerations^{17,21}. Before now, the Chorora Formation large mammal fauna was known from a fragmentary fossil assemblage recovered from a single locality, the ‘type site’, and included a primitive *Stegotetralodon* proboscidean, a large giraffid, a primitive hippopotamid, and a hipparionine equid^{17,18}. The latter is the most abundant taxon represented, and is considered the earliest known record of an equid in Africa with a best age-estimate of 10.7–10.1 Myr old²¹.

The Chorora fossils have been seen as a key assemblage for addressing both within-African and intercontinental mammalian evolutionary patterns²¹. This is because of Chorora’s intermediate geographic position between sub-Saharan Africa and Eurasia, as well as the paucity of African fauna spanning the time periods 12–7 Myr

ago. Despite such expectations, the fragmentary assemblages hitherto known have not realized the site’s potential.

We report here the first primate fossils recovered from the Chorora Formation. These were discovered during newly initiated systematic surveys from 2005 through to 2007 that aimed to locate palaeontological localities other than the relatively well-researched Chorora Formation type site. Primate remains were recovered at the Beticha locality, 3 km from the type site exposures, and include a macaque-sized cercopithecoid monkey (represented by a partial phalanx) and teeth of a large bodied hominoid primate. The primate fossils derive from the upper portions of the Chorora Formation, higher in the section than the type site assemblage. Their age is best considered as approximately 10.0 to 10.5 Myr old (Fig. 1, and Supplementary Information).

The large-bodied ape is represented by a canine and eight partial molars from at least three, and perhaps six or more, individuals (Fig. 1, and Supplementary Information). These teeth are collectively indistinguishable from modern gorilla subspecies in dental size and represented proportions (Supplementary Information). This modest sample nevertheless exhibits substantial size variation, with molars at both the largest and smallest end of the modern gorilla ranges of variation. The species also exhibits a suite of subtle but derived characteristic molar features that are shared with modern gorillas (see below). Functionally, they indicate incipient morphological specialization towards a relatively fibrous diet. Phylogenetically, these fossils represent the first Miocene ape species to be recognized as a strong candidate for membership in the modern gorilla clade.

Order Primates

Superfamily Hominoidea Gray, 1825

Chororapithecus gen. nov.

Chororapithecus abyssinicus sp. nov.

Etymology. The genus name is taken after Chorora, the name of the geological formation from which the fossils are derived; Chorora is the name of a local village approximately 8 km south of the Beticha locality. The species name is taken from Abyssinia, the former name of Ethiopia, in light of the geographical location of discovery at the junction area of the Main Ethiopian and Afar rift systems.

Holotype. CHO-BT 4 right upper M² (second molar).

Paratype. CHO-BT 3, left lower canine; CHO-BT 5, right upper M³; CHO-BT 6, right upper M³; CHO-BT 7, left lower M₃; CHO-BT 8, left lower M₁; CHO-BT 9, left lower molar fragment; CHO-BT 10, right lower molar fragment; and CHO-BT 11, right upper M³.

Diagnosis. A hominoid primate with molars of equivalent size to those of *Gorilla gorilla* subspecies that combine a tendency for enhanced shearing structures (relatively long mesial protocone crest of the upper molars and high trigonid crest of the lower molars) and

¹The University Museum, the University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. ²Department of Anthropology, National Museum of Nature and Science, Hyakunincho, Shinjuku-ku, Tokyo 169-0073, Japan. ³Division of Natural History, Hyogo Museum of Nature and Human Activities, Sanda, Hyogo 669-1546, Japan. ⁴Rift Valley Research Service, P.O. Box 5717, Addis Ababa, Ethiopia. ⁵Department of Archaeology and Paleontology, Authority for Research and Conservation of Cultural Heritage, Ministry of Culture and Tourism, P.O. Box 13247, Addis Ababa, Ethiopia.

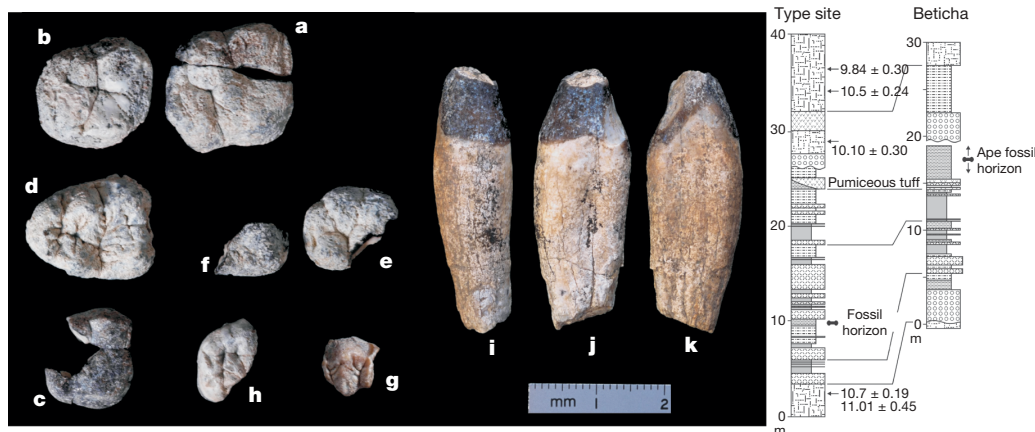


Figure 1 | *Chororapithecus abyssinicus* fossils and schematic stratigraphy. **a**, CHO-BT 4, mesiodistal dimension 19.1 mm, buccolingual dimension 18.0 mm. **b**, CHO-BT 5, mesiodistal dimension 16.9 mm, buccolingual dimension 17.5 mm. **c**, CHO-BT 6, mesiodistal dimension 13.2 mm, buccolingual dimension 15.0 mm. **d**, CHO-BT 7, mesiodistal dimension 19.1 mm, buccolingual dimension 15.9 mm. **e**, CHO-BT 8, buccolingual

dimension 13.2 mm. **f**, CHO-BT 9. **g**, CHO-BT 10. **h**, CHO-BT 11. **i–k**, CHO-BT 3 buccal, distal and mesial views, respectively, maximum diameter 13.9 mm, perpendicular diameter 10.7 mm. All dimensions except for those of CHO-BT 3 are estimated values corrected for minor damage. Simplified stratigraphy (type site section after ref. 20, radioisotopic age determinations from refs 17 and 20; see Supplementary Information for details).

relatively low cuspal topography with thick enamel at the ‘functional’ side cusps (lingual in upper and buccal in lower molars). The upper molars tend to be buccolingually narrow and mesiodistally elongate in shape, with a mesiobuccally extending mesial protocone crest. The epicrista is thin and buccolingually straight, creating a narrow mesial fovea located buccally. The high and continuous trigonid crest of the lower molars is situated distinctly above the level of a low mesial marginal ridge. The metaconid is moderately tall, forming a deep lingual notch in comparison with other large-bodied Miocene hominoids. In both upper and lower molars, cusp tips are relatively peripherally placed and cingular expressions are weak.

Differential diagnosis. *Chororapithecus abyssinicus* differs from *Gorilla gorilla* subspecies by having molar crown and cusps lower in height and with thicker enamel; upper molars with less distinct crista obliqua; and lower molars lacking buccolingual infolding of the buccal occlusal margin (that is, a lack of extremely elongated shearing crests). *C. abyssinicus* differs from all known modern and fossil hominoids (including the nearly contemporaneous African genus *Samburupithecus*), with the exception of *Gorilla*, *Rangwapithecus*, *Nyanzapithecus* and *Oreopithecus*, in enhanced shear between the mesial protocone crest and the distal occlusal slope of the metaconid. Furthermore, *C. abyssinicus* is different from *Samburupithecus*

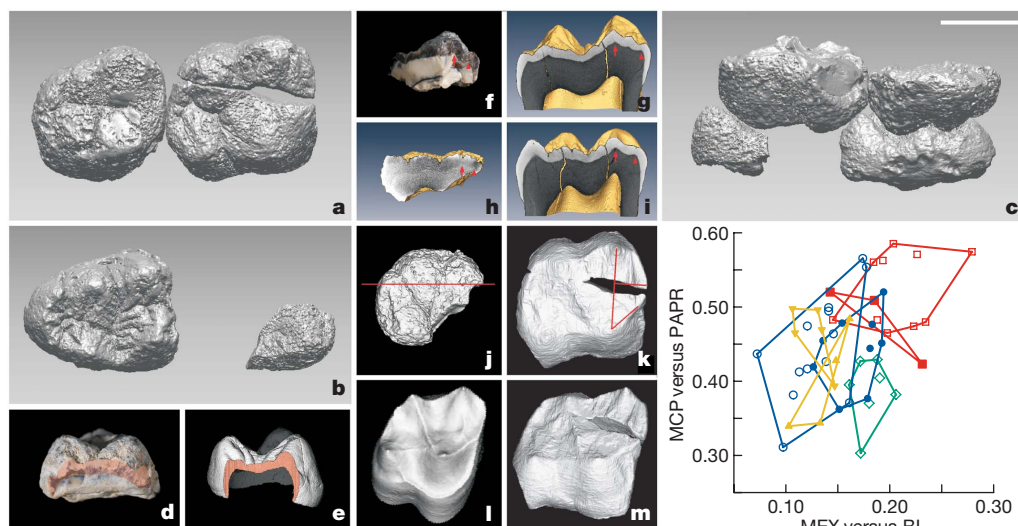


Figure 2 | Micro-CT-based morphological evaluations of the outer enamel surface and enamel-dentine junction. Three-dimensional rendered views of right upper M² and M³ (**a**), and left lower M₂ and M₃ (**b**). **c**, Lower molars mirror-imaged and placed in an occlusal relationship with the upper molars; lingual view, mesial to the left. Note the shearing relationship between mesiolingual protocone and the distal slope of the metaconid. Mesial view of *Chororapithecus* (**d**) and *Gorilla* (**e**) lower M₃s, with the mesial marginal ridge broken or cut off, showing a prominent trigonid crest. Natural section of *Chororapithecus* lower M₂ metaconid break (mirror-imaged, **f**), and equivalent micro-CT section of a *Gorilla* molar (**g**). **h**, *Chororapithecus* lower M₁ longitudinal section at mid-protoconid, its position as indicated in **j**, and equivalent section of a *Gorilla* molar (**i**). Note the prominent trigonid crest of the EDJ (left arrows) located distinctly above the level of the mesial marginal ridge (right arrowheads). Micro-CT visualization of EDJ surfaces of

Chororapithecus (**k**, **m**) and *Gorilla* (**l**) upper M²s. Occlusal view of *Chororapithecus* upper M² EDJ (**k**), showing the metrics evaluated in the right lower scatter plot. MCP-versus-PAPR is the distance between protocone tip and protoconule (or equivalent point) divided by paracone–protocone cusp tip distance, both measured on the EDJ. MEX-versus-BL is the projected length of the mesial fovea (measured from the paracone–protocone plane) divided by buccolingual crown breadth. Filled red squares, *Chororapithecus*; open squares, *Gorilla*; open diamonds, *Pongo*; open circles, *P. troglodytes*; filled circles, *P. paniscus*; filled triangles, *Australopithecus africanus* (pointing up) and *Australopithecus robustus* (pointing down). See the Supplementary Information for methodological details and comparative sample compositions. Note the linear to slightly concave mesial protocone crest of CHO-BT 4 resembling the *Gorilla* condition but with lower crown and cusp topography. Scale bar, 1 cm (common to all images).

because it has molars with a more open occlusal basin, weak expression of the anterolingual cingular complex, and a continuous crista obliqua (of M^2), as opposed to a crest incised by the longitudinal groove. *C. abyssinicus* differs from *Proconsul major*, *P. nyanzae* and 'Ugandapithecus' *gitonga*²² in its larger size, buccolingually narrower upper molars, more peripherally located molar cusps, resulting in a more open occlusal basin, thin and weak cingular expressions, and lack of tendency for a strong expression of the protoconule. Also, *C. abyssinicus* differs from *Ouranopithecus* in having a larger upper M^2 with a more mesiodistally elongate shape, better developed occlusal crests, and lack of lower M_3 posterior cusp development (hypoconulid and cusp 6 regions small); from *Ankarapithecus* and *Sivapithecus parvada* in its larger size, buccolingually narrower upper molars, and lack of tendency for a strong expression of the protoconule; and from *Gigantopithecus* in being smaller in size and possessing lower crowned molars, buccolingually narrow upper molars, better developed occlusal crests, and a lower canine with interlocking relationship with the upper canine. Other comparative remarks are given below and in the Supplementary Information.

The most distinctive features of the *Chororapithecus* dentition are the derived shearing structures seen in portions of its molars (Fig. 2), despite a generally low cuspal topography (the latter is apparently a primitive retention).

Examination of internal morphology by micro-computed tomography (micro-CT) demonstrates that these occlusal features were underlain by distinct enamel–dentine junction (EDJ) structure (Fig. 2). In particular, the straight to weakly concave mesial protocone crest seen in the EDJ of CHO-BT 4, -BT 5 and -BT 6 is gorilla-like, and is formed by a mesiobuccally located junction of the mesial protocone crest and mesial marginal ridge. Such spatial placements are best considered to be regulated by enamel-knot-related signalling patterns during early morphogenesis^{23,24}, and may be one of the underlying causes of the mesiodistally elongate upper molar shape generally characteristic of folivorous primate species. In the lower molars, the most distinctive EDJ topography occurs at the trigonid crest, the structural counterpart that occludes with the upper molar mesial protocone crest. The high trigonid EDJ crest is continuous between the metaconid and protoconid cusp tips (Fig. 2). Because recent experimental and quantitative genetic studies suggest significant degrees of morphogenetic independence between corresponding upper and lower molar structures^{25,26}, the presence of a functionally integral inter-jaw pattern of morphological expression, as seen in the *Chororapithecus* molars, suggests adaptation by natural selection, as opposed to chance emergence of neutral morphological minutia.

For the features in which *Chororapithecus* differs from gorillas, the gorilla condition is clearly the more derived towards advanced molar shear. In particular, *Chororapithecus* exhibits a much lower cusp relief and thicker molar enamel (Fig. 3). Maximum radial lateral thickness of the 'functional' side cusps (which form the main occlusal areas for crushing and grinding) of *Chororapithecus* appears to scale isometrically to that of *Australopithecus*, whereas enamel thickness of the opposite side cusps is equivalent to the intermediate conditions of *Pongo* and *Ardipithecus*. Enamel thickness is thought to probably have a significant developmental component directly related to steepness of EDJ topography²⁷. Scaled by basal crown area, enamel volume of *Chororapithecus* is intermediate between the *Gorilla* and *Australopithecus* conditions (Fig. 3). Thus, the higher crown and larger EDJ surface area of the gorilla molar would in part explain its thinner enamel coverage. Although the gorilla condition could potentially be attained from a thicker-enamelled ancestor, it is difficult to evaluate whether *Chororapithecus* itself could represent such an ancestral condition. This is because descriptions of EDJ topography and enamel thickness patterns among Miocene hominoids are limited. However, the thick enamel of the *Chororapithecus* 'functional' side cusps, and the extremely low EDJ topography seen in one of the upper molars (CHO-BT 4) indicate that *Chororapithecus* itself

is probably too specialized to represent a direct ancestral condition of the modern gorilla.

Thus, we consider *Chororapithecus* to exhibit a derived molar morphology indicative of incipient adaptation to a comparatively fibrous diet, as indicated by the emphasis of shear between the mesial protocone crest and the distal slope of the metaconid. At the same time, the low relief of the 'functional' side cusps endowed with thick enamel enable prolonged retention of an occlusal platform of rigid enamel (without exposing the more compliant dentine). This is reminiscent of three-dimensional enamel distribution patterns seen in *Pongo* and may be an adaptation for consuming hard and brittle food items²⁸. With such a unique combination of molar structures, the shearing crests of *Chororapithecus* may have functioned in the folding and pulverizing of fibrous foodstuff, rather than the actual cutting of fibrous materials by sharp, shearing crests, as in advanced folivores with higher cusps. So far, among the diversity of known African and Eurasian middle- to large-bodied Miocene apes, with the exception of *Oreopithecus* (which is even more uniquely specialized), we are not

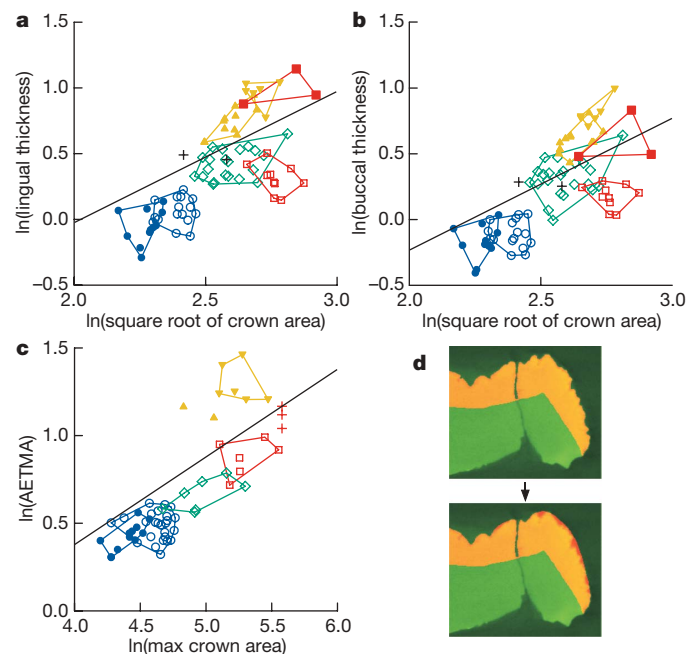


Figure 3 | Micro-CT-based evaluations of upper molar enamel thickness. Bivariate plots of ln(maximum radial enamel thickness) of the lateral crown faces of lingual 'functional' side cusps (a), and buccal 'non-functional' side cusps (b) by ln(crown area). The isometric line going through the mid-point of the *Ardipithecus ramidus* molars¹ ($n = 2$, crosses) is drawn for reference. Other symbols are as in Fig. 2. Reliable micro-CT-based estimates were obtained for three *Chororapithecus* upper molars as follows. CHO-BT 4: paracone, 1.64 mm; protocone, 2.53 mm; hypocone, 2.62 mm; CHO-BT 5: paracone, 2.32 mm; metacone, 2.28 mm; protocone, 3.28 mm; hypocone, 3.00 mm; CHO-BT 6: paracone, 1.62 mm, protocone, 2.41 mm. *Chororapithecus* shows a unique combination of thick enamel on the 'functional' side cusps, and intermediate thickness on the 'non-functional' side cusps. c, Log bivariate plot of AETMA (enamel volume/crown area) by crown area. Isometric line through the CHO-BT 4 estimate is shown for reference. The three crosses for CHO-BT 4 indicate the raw uncorrected enamel volume (lowest cross), correction for macro-damage (middle cross), and an additional 5% increase estimated for weathering and minor wear loss (upper cross). The degree of this correction in a representative section is depicted in d. Yellow, enamel; red, restored enamel crown. Other symbols are as in the previous plots. Note that with this measure of average enamel thickness, gorillas have enamel as 'thick' as the other apes; that is, despite the thinner enamel of the gorilla, total enamel volume available per crown area is comparable with other modern ape species, indicating broad equivalence in durability against abrasion. Note that *Chororapithecus* is intermediate between apes and *Australopithecus*. See Supplementary Information for the details of methodology and sample compositions.

aware of a single instance of a coherent set of derived features shared with the modern gorilla like those seen in *Chororapithecus*.

The similarities seen between the two genera raise the possibility that *Chororapithecus* is a Miocene member of the *Gorilla* clade. Alternatively, with its combination of thick enamel and distinct molar cresting pattern, *Chororapithecus* may represent a unique adaptation that is convergent with gorillas in molar structure and function. Although the evidence for phylogenetic affinity between *Chororapithecus* and *Gorilla* is inconclusive, it may be that the basal members of the gorilla clade shared large tooth size and incipiently enhanced molar shear as a part of an herbivorous diet that accompanied (presumed) larger body size. *Chororapithecus* may then represent one example of adaptational (and perhaps phyletic) differentiation within that clade.

Acceptance of *Chororapithecus* as a basal member of the gorilla clade would push back the gorilla species split to >10.5 Myr ago. Because this is a minimum date established from a meagre fossil record, the actual divergence would have predated this by an unknown time gap. From the currently available evidence, we consider that a species split of ~20 Myr ago for *Pongo*, 12 Myr ago for *Gorilla*, and 9 Myr ago for *Pan* are all probable estimates (see Supplementary Information). We consider that the early divergence hypothesis is congruent with both fossil and molecular data, and should be further evaluated using both sides of the evidence.

At Chorora, previous systematic field research at the type site has yielded a 'Hipparion'-dominated fauna, which was considered to indicate a significant representation of open habitats in possibly a mosaic environment¹⁷, an interpretation corroborated by recent isotopic and wear studies of 'Hipparion' teeth²¹. To the contrary, the Beticha locality is poorly fossiliferous, dominated by primates, and almost lacking in 'Hipparion' (Supplementary Information). The Beticha fossils occur at a higher stratigraphic level, within sediments formed by a braided river system with intermittent palaeosol formation (Supplementary Information). We hypothesize that this indicates a mosaic that includes a forested lake margin environment formed towards the end of sedimentation of the Chorora Formation palaeobasin, while episodic rifting was locally ceasing. Although the evidence is so far limited, this suggests that, in the changing landscape of the African middle to late Miocene, ape species were actually present in both what are now Kenya^{29,30} and Ethiopia. The available contextual evidence suggests that they predominantly occupied the more forested environments, and that the modern African ape and hominid clades probably emerged from a late Miocene African diversity of apes, which is only starting to be revealed. Palaeontologically, our results indicate that more focus is needed on comparatively poorly fossiliferous sediments that sample the habitats that were preferred by the Miocene apes.

Received 18 June; accepted 25 July 2007.

- White, T. D., Suwa, G. & Asfaw, B. *Australopithecus ramidus*, a new species of early hominid from Aramis, Ethiopia. *Nature* **371**, 306–312 (1994).
- Semaw, S. et al. Early Pliocene hominids from Gona, Ethiopia. *Nature* **433**, 301–305 (2005).
- Haile-Selassie, Y. Late Miocene hominids from the Middle Awash, Ethiopia. *Nature* **412**, 178–181 (2001).
- Haile-Selassie, Y., Suwa, G. & White, T. D. Late Miocene teeth from Middle Awash, Ethiopia, and early hominid dental evolution. *Science* **303**, 1503–1505 (2004).
- Senut, B. et al. First hominid from the Miocene (Lukeino Formation, Kenya). *C. R. Acad. Sci. (Paris)* **332**, 137–144 (2001).
- Brunet, M. et al. A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**, 145–151 (2002).
- Brunet, M. et al. New material of the earliest hominid from the Upper Miocene of Chad. *Nature* **434**, 752–755 (2005).
- White, T. D. et al. Asa Issie, Aramis and the origin of *Australopithecus*. *Nature* **440**, 883–889 (2006).
- Kimbel, W. H. et al. Was *Australopithecus anamensis* ancestral to *A. afarensis*? A case of anagenesis in the hominin fossil record. *J. Hum. Evol.* **51**, 134–152 (2006).
- Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. & Takahata, N. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl Acad. Sci. USA* **92**, 532–536 (1995).
- Chen, F.-C. & Li, W.-H. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* **68**, 444–456 (2001).
- Stauffer, R. L., Walker, A., Ryder, O. A., Lyons-Weiler, M. & Hedges, S. B. Human and ape molecular clocks and constraints on paleontological hypotheses. *J. Hered.* **92**, 469–474 (2001).
- Glazko, G. V. & Nei, M. Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* **20**, 424–434 (2003).
- Kumar, S., Filipowski, A., Swarna, V., Walker, A. & Hedges, S. B. Placing confidence limits on the molecular age of the human–chimpanzee divergence. *Proc. Natl Acad. Sci. USA* **102**, 18842–18847 (2005).
- Stewart, C.-B. & Disotell, T. R. Primate evolution — in and out of Africa. *Curr. Biol.* **8**, R582–R588 (1998).
- Begun, D. R. *Sivapithecus* is east and *Dryopithecus* is west, and never the twain shall meet. *Anthropol. Sci.* **113**, 53–64 (2005).
- Geraads, D., Alemseged, Z. & Bellon, H. The late Miocene mammalian fauna of Chorora, Awash basin, Ethiopia: systematics, biochronology and the ⁴⁰K–⁴⁰Ar ages of the associated volcanics. *Tertiary Res.* **21**, 113–122 (2002).
- Sickenberg, O. & Schönfeld, M. in *Afar Depression of Ethiopia* (eds Pilger, A. & Rösler, A.) 277–284 (Schweizerbart, Stuttgart, 1975).
- Kunz, K., Kreuzer, H. & Müller, P. in *Afar Depression of Ethiopia* (eds Pilger, A. & Rösler, A.) 370–374 (Schweizerbart, Stuttgart, 1975).
- Tiercelin, J.-J., Michaux, J. & Bandet, Y. Le Miocène supérieur du Sud de la Dépression de l'Afar, Éthiopie: sédiments, faunas, âges isotopiques. *Bull. Soc. Geol. Fr.* **21**, 255–258 (1979).
- Bernor, R. L., Kaiser, T. M. & Nelson, S. V. The oldest Ethiopian hipparion (*Equinae*, *Perissodactyla*) from Chorora: Systematics, paleodiet and paleoclimate. *Courier Forschungsinstit. Senckenberg* **246**, 213–226 (2004).
- Pickford, M. & Kunimatsu, Y. Catarrhines from the Middle Miocene (ca. 14.5 Ma) of Kipsaraman, Tugen Hills, Kenya. *Anthropol. Sci.* **113**, 189–224 (2005).
- Jernvall, J., Keränen, S. V. E. & Thesleff, I. Evolutionary modification of development in mammalian teeth: Quantifying gene expression patterns and topography. *Proc. Natl Acad. Sci. USA* **97**, 14444–14448 (2000).
- Kangas, A. T., Evans, A. R., Thesleff, I. & Jernvall, J. Nonindependence of mammalian dental characters. *Nature* **432**, 211–214 (2004).
- McCollum, M. & Sharpe, P. T. Evolution and development of teeth. *J. Anat.* **199**, 153–159 (2001).
- Hlusko, L. J. & Mahaney, M. C. Genetic contributions to expression of the baboon circular remnant. *Arch. Oral Biol.* **48**, 663–672 (2003).
- Suwa, G. & Kono, R. T. A micro-CT based study of linear enamel thickness in the mesial cusp section of human molars: reevaluation of methodology and assessment of within-tooth, serial, and individual variation. *Anthropol. Sci.* **113**, 273–289 (2005).
- Kono, R. T. Molar enamel thickness and distribution patterns in extant great apes and humans: new insights based on a 3-dimensional whole crown perspective. *Anthropol. Sci.* **112**, 121–146 (2004).
- Ishida, H. & Pickford, M. A new late Miocene hominoid from Kenya: *Samburupithecus kiptalami* gen. et sp. nov. *C. R. Acad. Sci. (Paris)* **325**, 823–829 (1997).
- Nakatsukasa, M. et al. Late Miocene primate fauna in Nakali, central Kenya. *Am. J. Phys. Anthropol.* **129** (suppl. 42), 136 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the National Science Foundation Revealing Hominid Origins Initiative (RHOI) and the Japan Society for the Promotion of Science for financial support of field and laboratory research. We thank W. Amgera and K. Kairente for field work; A. Ademassu for supporting laboratory work; J.-R. Boissier, M. Brunet, Y. Haile-Selassie, C. O. Lovejoy and T. White for observations and/or discussions; the staff of the Culture and Tourism Bureau Western Hararge Chiro Zone and the administration of the Mieso Woreda for fieldwork support; the Oromiya Culture and Tourism Bureau Addis Ababa for facilitation; and the Authority for Research and Conservation of Cultural Heritage and the National Museum of Ethiopia, Ministry of Culture and Tourism of Ethiopia for permissions and support. We thank the following institutions and staff for access to comparative materials: National Museum of Ethiopia; National Museums of Kenya; Transvaal Museum, South Africa; Royal Museum of Central Africa, Tervuren; Naturalis, Leiden; Cleveland Museum of Natural History; the University of California at Berkeley, Human Evolution Research Center; and the National Museum of Nature and Science, Tokyo, Department of Zoology.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.S. (suwa@um.u-tokyo.ac.jp).

Non-random coextinctions in phylogenetically structured mutualistic networks

Enrico L. Rezende¹, Jessica E. Lavabre¹, Paulo R. Guimarães Jr^{1,2}, Pedro Jordano¹ & Jordi Bascompte¹

The interactions between plants and their animal pollinators and seed dispersers have moulded much of Earth's biodiversity^{1–3}. Recently, it has been shown that these mutually beneficial interactions form complex networks with a well-defined architecture that may contribute to biodiversity persistence^{4–8}. Little is known, however, about which ecological and evolutionary processes generate these network patterns^{3,9}. Here we use phylogenetic methods^{10,11} to show that the phylogenetic relationships of species predict the number of interactions they exhibit in more than one-third of the networks, and the identity of the species with which they interact in about half of the networks. As a consequence of the phylogenetic effects on interaction patterns, simulated extinction events tend to trigger coextinction cascades of related species. This results in a non-random pruning of the evolutionary tree^{12,13} and a more pronounced loss of taxonomic diversity than expected in the absence of a phylogenetic signal. Our results emphasize how the simultaneous consideration of phylogenetic information and network architecture can contribute to our understanding of the structure and fate of species-rich communities.

Plant and animal species establish mutually beneficial interactions such as pollination and seed dispersal that can form complex networks of dependency. Recent work has characterized the architecture of mutualistic networks, with the ultimate goal of understanding their formation and maintenance and the coevolution of species within them. These networks are very heterogeneous (some species have a much larger number of interactions than expected by chance)⁴, and nested (specialists interact with proper subsets of the species that generalists interact with)⁵, and are built on weak and asymmetric links (for example, if a plant species depends strongly on an animal species, the dependence of the animal on that plant is much weaker)⁶. The next step on the road to understanding these networks is to disentangle the contribution of different evolutionary and ecological processes in generating their patterns^{3,9,14}.

Here we ask to what extent network architecture is associated with species phylogenetic relationships (Fig. 1), and whether coextinction cascades following a species disappearance⁷ involve phylogenetically related (that is, non-randomly sampled) species. The presence of a phylogenetic signal, where patterns of interactions between species can be partly explained by phylogenetic relatedness, would suggest that network patterns are partially dependent on past evolutionary history, and so cannot be exclusively explained by current ecological processes^{14–17}.

We compiled the largest data set of plant–animal mutualistic interactions, comprising 36 plant–pollinator and 23 plant–frugivore mutualistic networks spanning a broad geographic range (data sets are available as Supplementary Information). For each network, we reconstructed the phylogenies of the animals and plants (see Supplementary Methods). We then characterized two components of network architecture. First, we considered the number of interactions

per species, that is, species degree^{4,18}, and its quantitative extension, species strength. The strength of a plant species, for instance, is defined as the sum of dependences or interaction weights of the animal species on that plant⁶. These simple components of network architecture reflect the generalization level of a given species and its quantitative relevance in terms of how other species depend on it. Second, we considered the identity of each species' interactors (Methods). Both the number of interactions per species, and the identity of the species with which they interact have been identified as major determinants of network architecture and robustness^{7,19,20}. With the phylogenies and these components of network structure at hand, we applied phylogenetic statistical tools^{10,11} to characterize the extent to which closely related species tend to have similar patterns of interactions (Fig. 1).

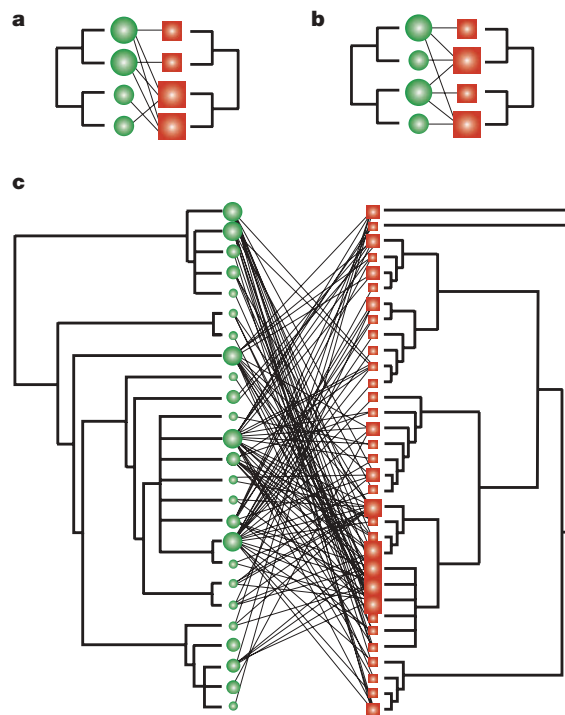


Figure 1 | A phylogenetic approach to mutualistic networks. We test to what extent the architecture of coevolutionary networks is associated with evolutionary history conveyed in the phylogenies of plants and animals. A plant (green circles) and an animal (red squares) are linked if the latter is a pollinator or a seed disperser of the former. Symbol size is proportional to its number of links. **a–c**, Examples where phylogeny accurately predicts the number of interactions (**a**), phylogeny does not predict the number of interactions (**b**), and the real correspondence in one network (**c**) (see NCOR in the Supplementary Methods).

¹Integrative Ecology Group, Estación Biológica de Doñana, CSIC, Apdo. 1056, E-41080 Sevilla, Spain. ²Instituto de Física 'Gleb Wataghin', UNICAMP, 13083-970, Campinas, São Paulo, Brazil.

Randomization tests (see Methods) suggest a significant phylogenetic signal in species degree in 24.8% of the data sets (26 of 105 phylogenies, one-tailed $P < 0.05$; Fig. 2a). Power analyses indicate that these estimates are highly conservative, suggesting that phylogenetic effects may be present in a larger proportion of the communities (Supplementary Figs 1 and 2). Also, the probability of detecting a signal in species degree seems to increase with phylogeny size (for example, 54.5% of the 22 phylogenies with more than 70 species showed significant signal; Supplementary Fig. 1). The negative association between the amount of phylogenetic signal estimated as K and phylogeny size (Fig. 2a) is probably an artefact. Even though the expectation of K is 1 under the null hypothesis of the true phylogeny, its lower bound decreases with the number of species in the phylogeny. Alternatively, Ornstein–Uhlenbeck branch-length transformation methods¹¹ (Methods) supported a significant signal in degree in 36.2% of the phylogenies where analyses converged (25 out of 69). P -values from randomization and branch-length transformation tests were highly positively correlated ($P < 0.001$), indicating that these two tests provide similar results. Thus these results show that phylogenetically related species have a similar number of interactions per species in at least one out of four phylogenies encompassing 39.0% of the networks.

Conversely, significant signal in species strength is present in only 2.6% of the phylogenies according to randomization tests (1 of 38; Fig. 2b), and in 20.8% of the phylogenies where branch-length transformation converged (5 of 24). The amount of phylogenetic signal for species strength was significantly lower than estimates for degree (paired t -test between log-transformed K values, $t_{37} = 1.806$, one-tailed $P = 0.039$; Fig. 2c). In spite of the significant positive correlation between degree and strength⁶, estimates of species strength may be subject to higher levels of uncertainty associated with proximate factors such as species abundance variability, changes associated with phenological sequences, and sampling errors that tend to decrease phylogenetic signal¹¹. This hypothesis could be tested by determining how estimates of species degree and strength vary over time in the same communities, the expectation being that species strength would show larger fluctuations than would degree. Alternatively, one could test whether the signal for strength increases after normalizing by species abundance. Nonetheless, our results suggest that species degree has stronger phylogenetic signal than strength.

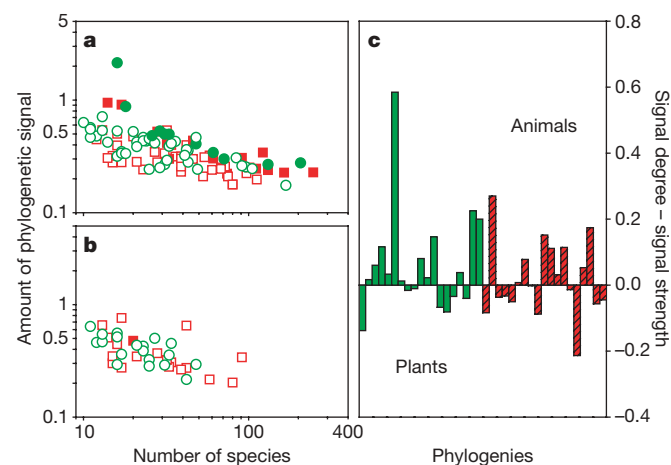


Figure 2 | Magnitude of phylogenetic signal on the number and strength of mutualistic interactions. Relationship between the magnitude of phylogenetic signal K and phylogeny size, estimated for species degree (a), and strength (b). Each data point represents a phylogeny: green circles for plants and red squares for animals. Solid symbols indicate statistically significant phylogenetic signals. c, Comparison of phylogenetic signal for species degree and strength. Green and dashed red bars correspond to plants and animals, respectively. Estimates obtained for degree were significantly higher.

Turning now to our second component of network architecture—which species interacts with which—we tested whether phylogenetic relatedness correlates with ecological similarity. The ecological similarity of any two species is defined as the number of species with which they both interact divided by the total number of species with which they interact (Methods). Phylogenetic and ecological distance matrices are positively and significantly correlated in 42.7% of the phylogenies (44 of 103, one-tailed Mantel test, $P < 0.05$). This means that phylogenetically related species tend to interact with a similar set of species. To determine whether this result is a consequence of the phylogenetic signal in degree reported above, we repeated these tests, controlling for differences in the number of interactions per species (partial Mantel test, see Methods). The results remained qualitatively similar. Partial Mantel correlations are significant in 46.6% of the phylogenies, supporting the idea that phylogeny is associated with the identity of the species' interactors after controlling for degree.

The association between phylogenetic resemblance and ecological similarity tends to be more common among animals: 60.8% of Mantel correlations between ecological and phylogenetic distance matrices were significant for animals, whereas 25.0% were significant for plants (Fig. 3). In addition, comparison of the Mantel coefficients Z for plants and animals indicated that animal phylogenies were more strongly associated with species interaction patterns than were plant phylogenies (paired t -test between log-transformed values, $t_{43} = 3.218$, one-tailed $P = 0.001$; Fig. 3). Further, these results are robust even when the number of animal and plant species is statistically controlled (analysis of covariance, ANCOVA, $F_{1,100} = 10.16$, $P = 0.02$). Results from partial Mantel tests controlling for degree were qualitatively similar (paired t -test, $t_{43} = 2.576$, one-tailed $P = 0.014$). It would be interesting to investigate multiple alternative hypotheses, such as differences in mobility and evolvability^{21–23}, to determine the cause of this difference.

Although network structure seems significantly more associated with animal phylogenies, structure may be driven by the evolutionary history of both plants and animals (7 of 44 communities), only plants or only animals (3 and 21 communities, respectively), or neither plants nor animals (the remaining 13 communities). This highlights the large variability across networks, indicated by the residual variation of K or Z after controlling for phylogeny size (Figs 2 and 3; Supplementary Table 1). Part of this variability is related to the taxonomic diversity of the plant lineages, though apparently not to the diversity of frugivores (fruit-eaters) or pollinators according to

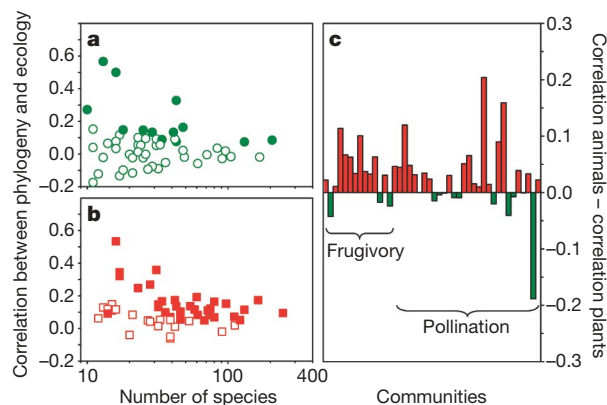


Figure 3 | Correlation between ecological similarity and phylogenetic relatedness. Results of regular Mantel tests correlating phylogenetic and ecological distance matrices plotted against phylogeny size, obtained for plants (a) and animals (b). Each data point corresponds to a phylogeny, and a solid symbol indicates a statistically significant correlation. c, Comparison between Mantel Z estimates obtained separately for plants and animals composing each network. Communities where interaction patterns are more associated with animal phylogenies are depicted in red; those more associated with plants in green. The phylogenetic structure of animals correlates significantly better with interaction matrices than that of plants.

several regression models (Methods). One plausible explanation for this result is that flower or fruit morphology is variable across the taxonomic groups examined, whereas phenotypic variation for frugivores and insects is more relevant at other taxonomic levels. Partitioning the phenotypic variance across taxonomic groups may address this hypothesis, and also clarify the mechanisms by which evolutionary history translates into patterns of interactions.

We have shown that phylogenetically related species tend to have similar roles in the network of interactions in almost half of the communities studied here. These effects are not a reflection of differences between major taxonomic groups only, because they also appear at finer scales of the phylogenies (Supplementary Fig. 3). In light of this, we tested whether simulated coextinctions⁷ involve taxonomically related species more often than expected by chance for cascades of identical size (Methods). This would result in a non-random pruning of the evolutionary tree^{12,13}. Simulations show that the rate of taxonomic diversity loss is higher than expected in the

absence of phylogenetic signal (Fig. 4). Although these effects may seem quantitatively small (partly owing to the averaging nature of the index), they can actually encompass the extinction of entire families or higher taxonomic groups from the community. The overall reduction in taxonomic diversity holds across communities, so that values falling below the null expectation are significantly more frequent than those above it ($\chi^2 = 50.7$, degrees of freedom, d.f. = 1, $P < 0.0001$; Fig. 4). Moreover, the contribution of phylogeny to species patterns of interaction correlates with the magnitude of taxonomic diversity loss across communities ($P < 0.05$ for parametric and non-parametric correlations; Fig. 4c). Therefore, communities in which species interactions have a strong phylogenetic component are more prone to have closely related species going coextinct following an extinction event. We conclude that the interaction between network and phylogenetic structures can ultimately result in non-random coextinction patterns.

Plant–animal mutualisms form heterogeneous, nested networks built on weak and asymmetric links among species, which may facilitate long-term species persistence^{6,7}. Our results provide evidence for the role of phylogenetic relationship as one determinant factor shaping these patterns: phylogeny partly accounts for species' propensities to interact in more than one-third of the networks, and the identity of the species with which they interact in about half of the networks. From a theoretical point of view, our results warrant the inclusion of evolutionary history into mechanistic models of network formation and maintenance²⁴. From a conservation perspective, our results show that cascading effects of coextinction may spread across taxonomically related species, further increasing the erosion of taxonomic diversity.

METHODS SUMMARY

The amount of phylogenetic signal in degree and strength was quantified with the K statistic, which is the fraction of the amount of signal of the real data set over that expected, assuming brownian motion and the same tree topology. Significance of phylogenetic signal was tested with randomization and branch-length transformation methods¹¹.

We used Mantel tests to compare phylogenetic distance matrices with matrices of ecological distance. Phylogenetic distance between pairs of plants (or animals) was estimated as the expected covariance of the trait between the two species^{11,25}. Ecological distance was calculated as $1 - S$, where S is the Jaccard index of similarity²⁶. We also performed partial Mantel tests controlling for the absolute difference in degree between two species.

Species removal simulations started from the most specialized to the most generalized species⁷. After an extinction cascade, we calculated the community taxonomic diversity as the average taxonomic distance between species²⁷. Path length weights between species increased the more distantly related they were taxonomically (that is, species of the same genus have a distance of 1 whereas species from different genera within the same family have a distance of 2, and so on). Thus, the higher the index, the more diverse the community. To calculate the decrease of taxonomic diversity of the real community with respect to the expected decrease in the absence of phylogenetic signal, we replicated the coextinction cascade 1,000 times after randomizing the taxonomic affiliation of species going coextinct. The taxonomic diversity relative to the null expectation was the ratio between real and null values, and the significance was estimated by counting how often the real value fell below the randomization results. The average rate of taxonomic loss per community was calculated as the slope of a linear regression with an intercept forced through 1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 March; accepted 25 May 2007.

1. Ehrlich, P. R. & Raven, P. H. Butterflies and plants: a study in coevolution. *Evol. Int. J. Org. Evol.* **18**, 586–608 (1964).
2. Thompson, J. N. *The Geographic Mosaic of Coevolution* (Univ. Chicago Press, Chicago, 2005).
3. Waser, N. M. & Ollerton, J. (eds) *Plant–Pollinator Interactions: from Specialization to Generalization* (Univ. Chicago Press, Chicago, 2006).
4. Jordano, P., Bascompte, J. & Olesen, J. M. Invariant properties in coevolutionary networks of plant–animal interactions. *Ecol. Lett.* **6**, 69–81 (2003).

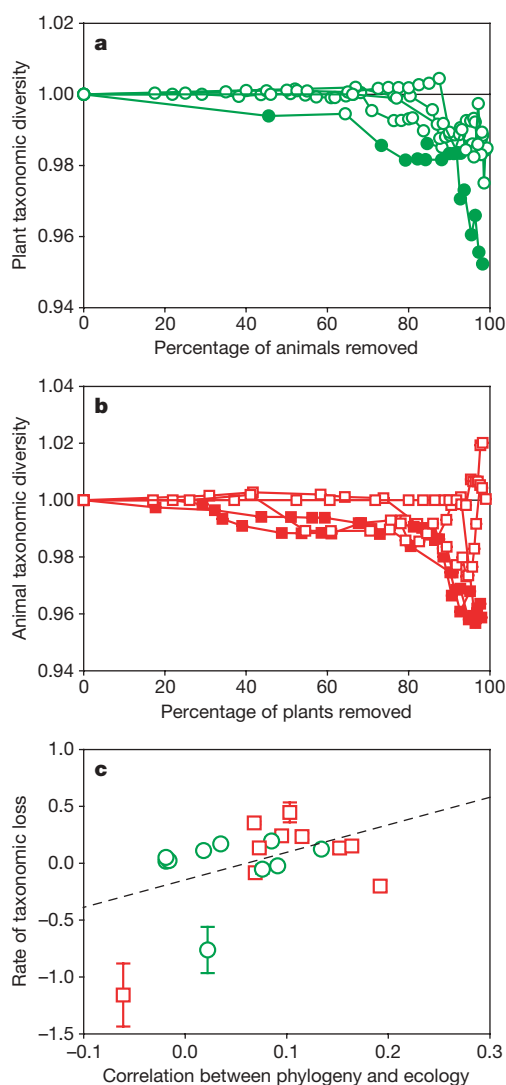


Figure 4 | Phylogenetic resemblance induces a higher loss of taxonomic diversity after species extinctions. Taxonomic diversity of plants (a) and animals (b) of five communities as a function of the number of extinct species, removed from most specialized to most generalized. Each symbol represents the removal step leading to the next generalization level. Taxonomic diversity is the ratio between indexes from the real communities over randomization results removing phylogenetic effects. Full symbols indicate taxonomic losses below 95% of the null model. c, Average (\pm s.e.m.) rate of taxonomic loss of ten communities regressed against the magnitude of phylogenetic effects.

5. Bascompte, J., Jordano, P., Melián, C. J. & Olesen, J. M. The nested assembly of plant–animal mutualistic networks. *Proc. Natl Acad. Sci. USA* **100**, 9383–9387 (2003).
6. Bascompte, J., Jordano, P. & Olesen, J. M. Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science* **312**, 431–433 (2006).
7. Memmott, J., Waser, N. M. & Price, M. V. Tolerance of pollination networks to species extinctions. *Proc. R. Soc. Lond. B* **271**, 2605–2611 (2004).
8. Vázquez, D. P. & Aizen, M. A. Asymmetric specialization: a pervasive feature of plant–pollinator interactions. *Ecology* **85**, 1251–1257 (2004).
9. Thompson, J. N. Mutualistic webs of species. *Science* **312**, 372–373 (2006).
10. Freckleton, R. P., Harvey, P. H. & Pagel, M. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* **160**, 712–726 (2002).
11. Blomberg, S. P., Garland, T. Jr & Ives, A. R. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evol. Int. J. Org. Evol.* **57**, 717–745 (2003).
12. Purvis, A., Agapow, P.-M., Gittleman, J. J. & Mace, G. M. Nonrandom extinction and the loss of evolutionary history. *Science* **288**, 328–330 (2000).
13. Heard, S. B. & Mooers, A. Ø. Phylogenetically patterned speciation rates and extinction risks change the loss of evolutionary history during extinctions. *Proc. R. Soc. Lond. B* **267**, 613–620 (2000).
14. Ives, A. R. & Godfray, C. J. Phylogenetic analysis of trophic associations. *Am. Nat.* **168**, E1–E14 (2006).
15. Armbruster, W. S. Phylogeny and the evolution of plant–animal interactions. *Bioscience* **42**, 12–20 (1992).
16. Herrera, C. M. Historical effects and sorting processes as explanations for contemporary ecological patterns: character syndromes in Mediterranean woody plants. *Am. Nat.* **140**, 421–446 (1992).
17. Jordano, P. Angiosperm fleshy fruits and seed dispersers: a comparative analysis of adaptation and constraints in plant–animal interactions. *Am. Nat.* **145**, 163–191 (1995).
18. Waser, N. M., Chittka, L., Price, M. V., Williams, N. M. & Ollerton, J. Generalization in pollination systems, and why it matters. *Ecology* **77**, 1043–1060 (1996).
19. Albert, R., Barabási, A.-L. & Jeong, H. The Internet's Achilles' Heel: Error and attack tolerance in complex networks. *Nature* **406**, 378–382 (2000).
20. Montoya, J. M., Pimm, S. L. & Solé, R. V. Ecological networks and their fragility. *Nature* **442**, 259–264 (2006).
21. Wheelwright, N. T. Fruit size in a tropical tree species—variation, preference by birds, and heritability. *Vegetatio* **108**, 163–174 (1993).
22. Fenster, C. B., Armbruster, W. S., Wilson, P., Dudash, M. R. & Thompson, J. D. Pollination syndromes and floral specialization. *Annu. Rev. Ecol. Syst.* **35**, 375–403 (2004).
23. Bronstein, J. L., Alarcón, R. & Geber, M. The evolution of plant–insect mutualisms. *New Phytol.* **172**, 412–428 (2006).
24. Cattin, M.-F., Bersier, L.-F., Banasek-Richter, C., Baltensperger, R. & Gabriel, J.-P. Phylogenetic constraints and adaptation explain food-web structure. *Nature* **427**, 835–839 (2004).
25. Garland, T. A. Bennett, F. & Rezende, E. L. Phylogenetic approaches in comparative physiology. *J. Exp. Biol.* **208**, 3015–3035 (2005).
26. Legendre, P. & Legendre, L. *Numerical Ecology* 2nd edn, 254–256 (Elsevier, Amsterdam, 1998).
27. Clarke, K. R. & Warwick, R. M. A taxonomic distinctness index and its statistical properties. *J. Appl. Ecol.* **35**, 523–531 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Armbruster, P. Buston, M. A. Fortuna, P. Guimarães, M. Helmus, A. Ives, J. Olesen, D. Posada, A. G. Sáez, J. N. Thompson and N. Waser for comments on a previous draft. This work was funded by the European Heads of Research Councils, the European Science Foundation, and the EC Sixth Framework Programme through a EURI (European Young Investigator) Award (J.B.), by the Spanish Ministry of Education and Science (P.J. and J.B.), by the Junta de Andalucía (P.J. and J.B.), and by the Fundação de Amparo à Pesquisa do Estado de São Paulo (P.R.G.).

Author Contributions E.L.R. performed all the analysis and compiled the phylogenies jointly with J.E.L. P.R.G. performed the extinction simulations. P.J. and J.B. designed the study and compiled the interaction matrices. E.L.R. and J.B. wrote a first version of the manuscript, and all authors contributed to the final draft.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.B. (bascompte@ebd.csic.es).

METHODS

Database. We compiled 59 qualitative mutualistic networks (36 plant–pollinator and 23 plant–frugivore webs) describing the presence or absence of interactions. The data set encompasses Mediterranean, tropical, temperate, subtropical and Arctic communities from all continents except mainland Asia and Antarctica. Of these networks, 22 (9 for pollination and 13 for frugivory) are quantitative, describing the strength of each interaction or pairwise dependence (see data sets in Supplementary Information). From these networks, we calculated species degree and species strength for plants and animals separately, and recorded the taxonomic affiliation of species forming each community.

Phylogenies. We assembled one animal and one plant phylogeny for each community. Phylogenies are based primarily on molecular data, with a few species included according to taxonomic information (Supplementary Methods). Phylogenies with less than ten species or with too many unresolved nodes were not included in analyses. This resulted in 105 phylogenies for the following groups: 35 insect phylogenies (Class Insecta; all pollinators), 18 bird phylogenies (Class Aves; all frugivores), and 52 angiosperm phylogenies (Infraphylum Angiospermae; 33 belonging to plant–pollinator and 19 to plant–frugivore networks).

Phylogenetic statistical methods. We tested for the presence of phylogenetic signal on species degree and strength with randomization and branch-length-transformation tests¹¹. These methods test whether species attributes are significantly associated with phylogeny, using randomization or maximum-likelihood procedures. Branch-length transformation tests were performed assuming the Ornstein–Uhlenbeck model of stabilizing selection and a model in which character evolution can accelerate or decelerate (ACDC)¹¹. Because ACDC models did not converge in most cases, here we discuss results from the Ornstein–Uhlenbeck model (all analyses are included in Supplementary Material for completeness). Although these techniques provide similar information about the presence of a phylogenetic signal, applying both methods can be useful in determining how robust our results are and in overcoming limitations inherent to each statistical test (Supplementary Figs 1 and 2). The amount of phylogenetic signal was quantified with the *K* statistic (Fig. 2), which is the fraction of the amount of signal of the real data set over that expected, assuming brownian motion and the same tree topology.

We used Mantel tests to compare phylogenetic distance matrices with matrices of ecological distances between species. Phylogenetic distance between pairs of plants (or animals) was estimated as the expected covariance of the trait between the two species^{11,25}. Ecological distance was calculated as $1 - S$, where *S* is the Jaccard index of similarity obtained from qualitative interaction matrices²⁶. The similarity between two species *i* and *j* is defined as $S(i, j) = a/(a + b + c)$, where *a*, *b* and *c* represent the number of shared interacting

species, the number of interactions specific to species *i*, and the number of interactions exclusive to species *j*, respectively.

Because differences in degree affect Jaccard estimates, we also performed partial Mantel tests controlling for degree (the pairwise distance in degree was calculated as the absolute difference in degree between two species; Supplementary Methods). Hence, this partial test can discern whether phylogeny strictly affects the species with which species interact, independently of the total number of interactions of each species. When necessary, estimates were log-transformed to improve normality (or log-value + 1 for statistics varying between −1 and 1, as Mantel's *Z*).

According to regression models controlling for phylogeny size, community size, and number of interactions, phylogenetic signal was similar for frugivory and pollination networks ($P > 0.28$ for *K* and for Mantel's regular and partial *Z*), hence results were pooled.

Taxonomic diversity and coextinction simulations. As a surrogate for phylogenetic diversity, we estimated taxonomic diversity of plants and animals in the largest available phylogenies (23 plant and 27 pollinator phylogenies with more than 30 species, and 15 bird phylogenies with more than 15 species; see Supplementary Methods). The mean taxonomic distance between all species was employed as an index of taxonomic diversity²⁷ in subsequent regressions.

Extinction cascades were simulated for the ten largest communities (all having more than 40 animal and plant species) with available taxonomic affiliation, following ref. 7. After one species is removed, species left without any interaction go coextinct. Species removal started from the most specialized (least-linked) to the most generalized (most-linked) species, which was proposed as a more plausible extinction sequence because specialist species tend to be less abundant than generalists^{4,7,8}. In spite of several assumptions implicit in the model (for example, all plants require animals for reproduction, and species cannot adapt to new resources), this approach provides the first reasonable attempt to study coextinction patterns in phylogenetically structured networks⁷. Although these assumptions may affect the total number of species going extinct⁷, our comparisons involve coextinction cascades of the same size, with and without phylogenetic signal.

After an extinction cascade, we calculated the decrease of taxonomic diversity of the real community respect to the expected decrease in the absence of phylogenetic signal. This was done by replicating the coextinction cascade after randomizing the taxonomic affiliation of species going coextinct (that is, nodes remain unchanged but their 'name tags' are shuffled). This null model removes effects of phylogenetic relatedness¹¹ controlling for network structure and species number. The relative taxonomic diversity is the ratio between real and null values, and the average rate of taxonomic loss per community is the slope of a linear regression with an intercept forced through 1 (that is, real values and the null expectation are equal when no species are removed).

Regulation of IgA production by naturally occurring TNF/iNOS-producing dendritic cells

Hiroiyuki Tezuka¹, Yukiko Abe¹, Makoto Iwata², Hajime Takeuchi², Hiromichi Ishikawa³, Masayuki Matsushita⁴, Tetsuo Shiohara⁵, Shizuo Akira⁶ & Toshiaki Ohteki¹

Immunoglobulin-A has an irreplaceable role in the mucosal defence against infectious microbes^{1–6}. In human and mouse, IgA-producing plasma cells comprise ~20% of total plasma cells of peripheral lymphoid tissues, whereas more than 80% of plasma cells produce IgA in mucosa-associated lymphoid tissues (MALT)^{1–6}. One of the most biologically important and long-standing questions in immunology is why this 'biased' IgA synthesis takes place in the MALT but not other lymphoid organs. Here we show that IgA class-switch recombination (CSR) is impaired in inducible-nitric-oxide-synthase-deficient (*iNOS*^{−/−}; gene also called *Nos2*) mice. iNOS regulates the T-cell-dependent IgA CSR through expression of transforming growth factor- β receptor, and the T-cell-independent IgA CSR through production of a proliferation-inducing ligand (APRIL, also called *Tnfsf13*) and a B-cell-activating factor of the tumour necrosis factor (TNF) family (BAFF, also called *Tnfsf13b*). Notably, iNOS is preferentially expressed in MALT dendritic cells in response to the recognition of commensal bacteria by toll-like receptor. Furthermore, adoptive transfer of iNOS⁺ dendritic cells rescues IgA production in *iNOS*^{−/−} mice. Further analysis revealed that the MALT dendritic cells are a TNF- α /iNOS-producing dendritic-cell subset, originally identified in mice infected with *Listeria monocytogenes*^{7,8}. The presence of a naturally occurring TNF- α /iNOS-producing dendritic-cell subset may explain the predominance of IgA production in the MALT, critical for gut homeostasis.

IgA is the most abundant immunoglobulin in the body and has a critical role in mucosal immunity^{1,2}. Secretory IgA is constitutively released into the lumen via the transport system of the intestinal epithelia³; it prevents invading pathogens from binding to mucosal epithelial cells and neutralizes their toxins^{4–6}. Natural secretory IgA is induced by constant antigenic stimulation by the intestinal commensal bacteria, and dendritic cells have an important function in this process^{9,10}. IgA-specific CSR occurs in the constant region of the immunoglobulin heavy chain (C_H), from C μ to C α , in naive B cells, and is accomplished by T-cell-dependent and T-cell-independent CSR⁵. T-cell-dependent CSR is induced by TGF- β 1 and CD40L expressed on activated T cells. On the other hand, two dendritic-cell-derived cytokines, APRIL and BAFF, directly mediate T-cell-independent CSR^{11,12}. Notably, the IgA CSR mainly takes place in the MALT, particularly in Peyer's patches and lamina propria, then IgA⁺ B cells re-circulate through other lymphoid organs and the bloodstream^{5,13,14}. In this context, one of the more enigmatic questions is why IgA CSR and IgA secretion take place largely in the MALT but not other lymphoid organs.

iNOS is upregulated by bacterial components and modulates immune responses¹⁵. However, little is known about the role of

iNOS in immunoglobulin synthesis. Under steady-state conditions, the serum IgA and IgG2b levels in *iNOS*^{−/−} mice were significantly lower compared with wild-type mice, whereas the serum levels of other immunoglobulin subclasses were unaffected in *iNOS*^{−/−} mice (Fig. 1a). *iNOS*^{−/−} mice also showed impaired secretory IgA production in both intestinal contents and faecal pellets (Fig. 1a). Notably, the impaired IgA production strictly correlates with the lack of the *iNOS* gene even in F₂ ((wild type \times *iNOS*^{−/−}) F₁ \times *iNOS*^{−/−}) offspring (Supplementary Fig. 1a), and the level of IgA in the sera and intestinal contents of wild-type mice treated with selective iNOS inhibitors (aminoguanidine (AMG) and L-N⁶-(1-iminoethyl)-lysine (L-NIL)) and a NO scavenger (carboxy-PTIO (C-PTIO)) was significantly lower than in untreated wild-type mice (Fig. 1b). In contrast, the number of intestinal T-cell-receptor (TCR)- $\gamma\delta$ ⁺ cells and the expression level of polymeric immunoglobulin receptor (pIgR) messenger RNA in intestinal epithelial cells—critical regulators for the transportation of secretory IgA into the lumen^{16,17}—were unaffected in *iNOS*^{−/−} mice (Supplementary Fig. 1b, c). As expected from these observations, the number of post-CSR IgA⁺ B cells (IgA⁺ B220⁺) and IgA⁺ plasma cells (IgA⁺ B220[−]), and the *in vitro* production of IgA of these cells in the absence or presence of interleukin (IL)-5, were significantly reduced in *iNOS*^{−/−} mice (Fig. 1c, f and Supplementary Fig. 1d) and wild-type mice treated with AMG, L-NIL and C-PTIO (Fig. 1d, e). This effect was probably not due to the impaired trafficking and survival of IgA⁺ B cells (Supplementary Fig. 1e), and the expression level of IL-5 receptor and IL-5-dependent proliferation were unaffected in B cells of *iNOS*^{−/−} mice (Supplementary Fig. 1f). Notably, impaired ovalbumin (OVA)-specific IgA response was also apparent in *iNOS*^{−/−} mice (Fig. 1g). Thus, iNOS-derived NO is required for IgA synthesis *in vivo*.

IgA CSR is mediated by both T-cell-dependent and T-cell-independent pathways⁵. To test whether the T-cell-dependent IgA CSR is impaired in *iNOS*^{−/−} B cells, naive IgD⁺ B cells of wild-type and *iNOS*^{−/−} mice were stimulated with TGF- β 1 and anti-CD40 monoclonal antibody, an inducer for IgA CSR⁵. Additional stimulation, such as anti-CD40 plus interferon- γ and anti-CD40 plus IL-4, was also included to induce IgG2c and IgG1/IgE CSR, respectively. Not only immunoglobulin production but also the expression of activation-induced cytidine deaminase (AID, an essential enzyme for CSR¹⁸) and α -germinal transcript (α -GT, a prerequisite for C μ to C α CSR) were impaired in *iNOS*^{−/−} B cells under IgA-CSR-inducing conditions (Fig. 2a–c), indicating that the IgA CSR is impaired in *iNOS*^{−/−} B cells. The TGF- β receptor, an essential component of T-cell-dependent IgA CSR, is composed of TGF- β RI and TGF- β RII^{19,20}. Of note, the expression of TGF- β RII, but not TGF- β RI, in naive B cells prepared from the indicated organs of *iNOS*^{−/−}

¹Department of Immunology, Akita University Graduate School of Medicine, Akita 010-8543, Japan. ²Faculty of Pharmaceutical Sciences at Kagawa Campus, Tokushima Bunri University, Kagawa 769-2193, Japan. ³Department of Microbiology and Immunology, Keio University School of Medicine, Tokyo 160-8582, Japan. ⁴Mitsubishi Kagaku Institute of Life Sciences, Tokyo 194-8511, Japan. ⁵Department of Dermatology, Kyorin University School of Medicine, Tokyo 181-8611, Japan. ⁶Department of Host Defense, Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan.

mice was much lower compared with that of wild-type mice (Fig. 2d, e). Thus, the TGF- β 1- and anti-CD40-mediated induction of downstream molecules of the TGF- β receptor, such as Smad3 (ref. 21) and Runx3 (ref. 22), was decreased in *iNOS*^{-/-} B cells (Fig. 2f). In contrast, neither CD40 expression nor CD40-dependent proliferation was affected in these cells (Supplementary Fig. 2a). Notably, the B-cell expression of TGF- β RII was restored when naive *iNOS*^{-/-} B cells were cultured with NONOate, a NO donor, in a phosphatidylinositol-3 kinase/AKT-dependent manner *in vitro* (Supplementary Fig. 2b) and transferred into μ Mt (known as Igh-6) mice (Fig. 2g). The latter was coincident with the restoration of serum IgA level (Fig. 2h), indicating that *iNOS*^{-/-} B cells retain the ability to secrete IgA in *iNOS*^{+/+} host. These results demonstrate that the critical role for iNOS in the T-cell-dependent IgA CSR is to regulate the B-cell expression of TGF- β RII.

To examine further whether *iNOS*^{-/-} T cells are functional in T-cell-dependent IgA production *in vivo*, T cells isolated from wild-type and *iNOS*^{-/-} mice were transferred into T-cell-deficient *Tcrb*^{-/-}*Tcrd*^{-/-} mice, in which systemic and secretory IgA production was impaired owing to the lack of T-cell-dependent CSR (Supplementary Fig. 3a)¹⁰. With time, the serum and intestinal content IgA production was restored in *Tcrb*^{-/-}*Tcrd*^{-/-} mice that received either wild-type or *iNOS*^{-/-} T cells, at a comparable level (Supplementary Fig. 3b). Furthermore, in the presence of TGF- β 1, naive wild-type B cells co-cultured with activated *iNOS*^{-/-} T cells produced IgA at a similar level to those co-cultured with activated wild-type T cells (Supplementary Fig. 3c). These

results indicate that *iNOS*^{-/-} T cells are functional in T-cell-dependent IgA synthesis.

We next examined T-cell-independent IgA production, which is mediated by dendritic-cell-derived cytokines APRIL and BAFF, as well as B-1 cells¹⁰⁻¹². Of note, iNOS was co-expressed with APRIL and BAFF in Peyer's patches and lamina propria dendritic cells of wild-type mice, whereas the expression level of these cytokines was decreased in the equivalent dendritic cells of *iNOS*^{-/-} mice (Fig. 3a, b and Supplementary Fig. 4a). In this context, on CpG stimulation, expression of APRIL and BAFF was elevated in wild-type dendritic cells but not in *iNOS*^{-/-} dendritic cells nor wild-type dendritic cells treated with L-NIL (Supplementary Fig. 4b). In contrast, expression of transmembrane activator and CAML interactor (TACI), a receptor for APRIL and BAFF, was unaffected in *iNOS*^{-/-} B cells (Supplementary Fig. 4c). As expected from these findings, IgA production from *iNOS*^{-/-} B cells was impaired when co-cultured with Peyer's patches and lamina propria dendritic cells of *iNOS*^{-/-} mice, whereas it was significantly restored when co-cultured with the equivalent dendritic cells of wild-type mice or those of *iNOS*^{-/-} mice in the presence of APRIL (Fig. 3c, d). The T-cell-independent IgA production was efficiently blocked in the presence of BCMA- and TACI-Ig (Supplementary Fig. 4d). In this context, *iNOS*^{-/-} dendritic-cell expression of APRIL was substantially restored when cultured with NONOate in a NFAT/NF- κ B-dependent manner, whereas expression of TGF- β RII by *iNOS*^{-/-} B cells remained impaired even after APRIL stimulation (Supplementary Fig. 4e, f). In addition, the level of APRIL-induced IgA synthesis was unaffected in the presence of

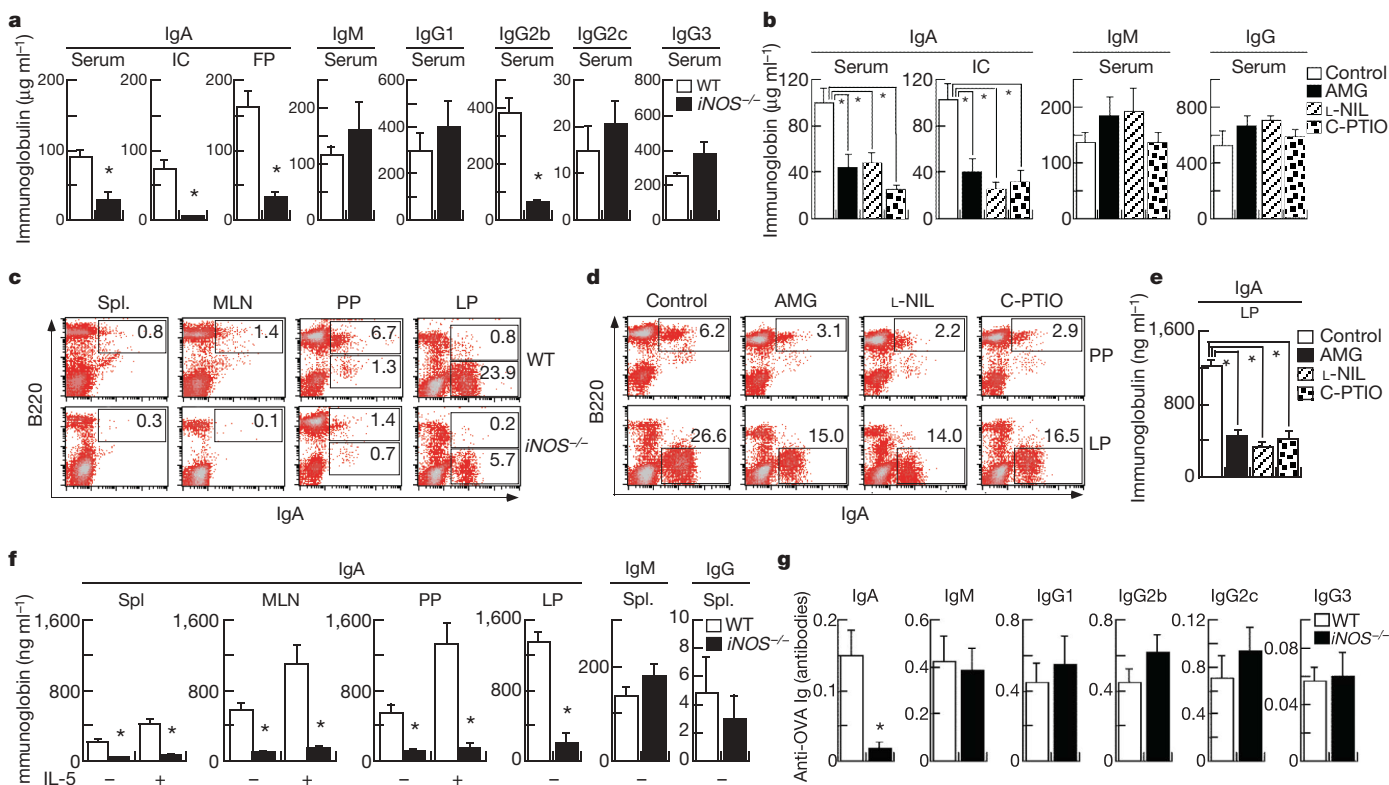


Figure 1 | IgA reduction in *iNOS*^{-/-} mice. **a, b**, Samples from the sera, intestinal content (IC) and faecal pellets (FP) of control wild-type and *iNOS*^{-/-} mice (**a**), and wild-type mice given 2.5% (w/v) AMG for 7 days, 4.5 mM L-NIL for 14 days, and 3 mg day⁻¹ C-PTIO for 14 days (**b**), were tested for immunoglobulin levels by ELISA. Each group includes three to four mice and data are representative of three experiments. **c, d**, Mononuclear cells from the spleen (Spl.), mesenteric lymph node (MLN), Peyer's patches (PP) and lamina propria (LP) cells of indicated mice were stained for IgA and B220. Numbers in each rectangle indicate the proportion of the cells. Data are representative of three experiments. **e**, Lamina propria cells of AMG-, L-NIL- and C-PTIO-treated wild-type mice were cultured for 3 days. **f**, B cells

from the spleen, mesenteric lymph node and Peyer's patches, and total LP cells, of indicated mice were cultured in the absence or presence of 100 ng ml⁻¹ IL-5 for 7 days. Immunoglobulin level in the culture medium is shown in **e** and **f**. Data in **e, f**, show mean \pm s.d. of triplicate cultures and are representative of five experiments. **g**, Mice were immunized with 100 μ g OVA, boosted with 50 μ g OVA 21 days after immunization, and OVA-specific immunoglobulin levels were determined 10 days later by ELISA. Each group includes three mice and data are representative of two experiments. Asterisks in relevant panels indicate unpaired Student's *t*-test; *P* value is <0.05. Error bars represent s.d.

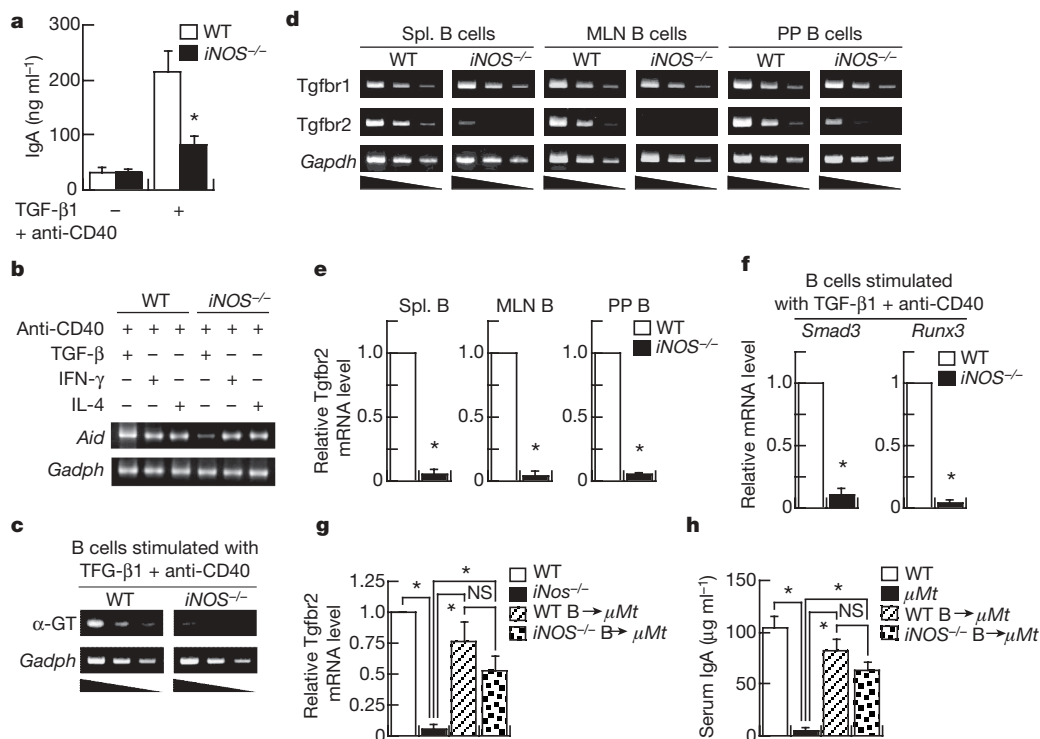


Figure 2 | Impaired T-cell-dependent IgA CSR in *iNOS*^{-/-} mice. Naive IgD⁺ B cells isolated from the spleen of wild-type and *iNOS*^{-/-} mice were cultured with the indicated stimuli for 7 days (a) or 3 days (b, c, f). The cytokines and monoclonal antibody used in this experiment were: 1 ng ml⁻¹ TGF-β1, 100 ng ml⁻¹ interferon-γ, 100 ng ml⁻¹ IL-4, 2 μg ml⁻¹ anti-CD40 monoclonal antibody. a, IgA production by stimulated B cells was determined by ELISA. Data show mean ± s.d. of triplicate cultures. b, c, *Aid* (also called *Aicda*) and α-GT (an intermediate germinal transcript for C_μ to C_α CSR) mRNA expression in stimulated B cells was examined by RT-PCR. d, e, TGF-βRI and

TGF-βRII expression levels (*Tgfb1* and *Tgfb2*, respectively) in naive B cells of the indicated tissues were evaluated by RT-PCR (d) and real-time PCR (e, f). Expression of *Smad3* and *Runx3* mRNA in stimulated B cells was assessed by real-time PCR. g, h, Wild-type and *iNOS*^{-/-} naive IgD⁺ B cells were transferred into *μMt*^{-/-} mice. Three weeks after the transfer, naive B cells were tested for TGF-βRII expression by real-time PCR (g) and serum IgA level was determined by ELISA (h). NS, not significant. Data are representative of five experiments (b–d) or show mean ± s.d. of three samples (e–h). Asterisks in relevant panels indicate unpaired Student's *t*-test; *P* value is <0.05.

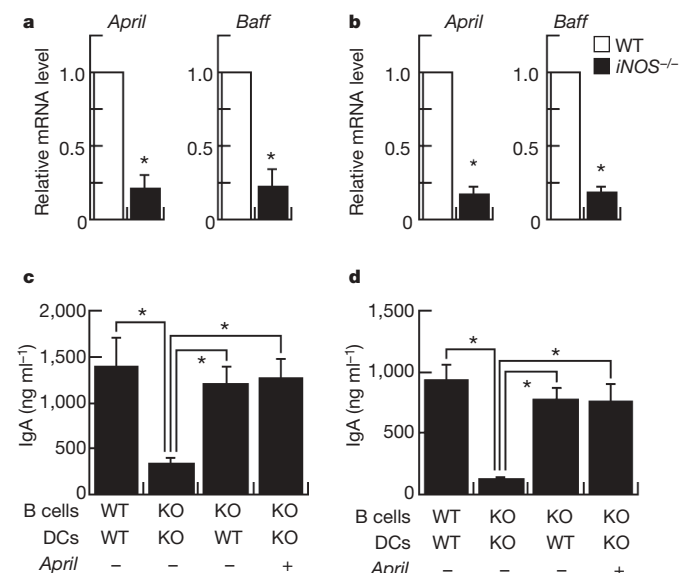


Figure 3 | Impaired T-cell-independent CSR in *iNOS*^{-/-} mice. a, b, Peyer's patches (a) and lamina propria (b) dendritic cells of wild-type and *iNOS*^{-/-} mice were examined for *April* and *Baff* expression by real-time PCR. Data show mean ± s.d. of three samples. c, d, Splenic IgD⁺ B cells of wild-type and *iNOS*^{-/-} (KO) mice were cultured with Peyer's patches (c) and lamina propria (d) dendritic cells (DCs) of wild-type and *iNOS*^{-/-} (KO) mice in the presence or absence of 100 ng ml⁻¹ APRIL for 7 days, and the IgA level was evaluated by ELISA. Data show mean ± s.d. of triplicate cultures and are representative of five experiments. Asterisks in relevant panels indicate unpaired Student's *t*-test; *P* value is <0.05.

anti-TGF-β neutralizing antibody or TGF-β1 (Supplementary Fig. 4g), showing that APRIL regulates IgA synthesis independently of TGF-β/TGF-βR in mice. Furthermore, the number and *in vitro* IgA production of B-1 cells were reduced in *iNOS*^{-/-} mice (Supplementary Fig. 4h, i).

These results indicate that iNOS might be expressed in the dendritic cells residing in the sites where IgA CSR predominates *in vivo*. Notably, iNOS was preferentially expressed in CD11c^{lo} dendritic cells in the MALT. The number of iNOS⁺ dendritic cells was significantly higher in the lamina propria than in the Peyer's patches and mesenteric lymph node, and negligible in the spleen, of wild-type mice (Fig. 4a), which correlates well with the IgA production in each site (Fig. 1f). To identify the induction mechanism of iNOS⁺ dendritic cells in the MALT under steady-state conditions, we further analysed *Myd88*^{-/-}, germ-free and *Tlr2*^{-/-} *Tlr4*^{-/-} *Tlr9*^{-/-} mice. The number of iNOS⁺ dendritic cells was much reduced in the MALT of these mice (Fig. 4a and Supplementary Fig. 5a), suggesting that this dendritic cell subset is induced by TLR-dependent recognition of commensal bacteria. The lamina propria dendritic cells also expressed gaseous NO, suggesting that NO is produced by iNOS⁺ dendritic cells (Fig. 4b). The iNOS⁺ dendritic cells were TNF-α⁺ CD11b⁺ MHC class II⁺ CD80⁺ CD86^{lo} Ly6C⁺ Ly6G⁻ Gr-1⁺ Mac3⁺, revealing that they are presumably a naturally occurring TNF-α/iNOS-producing dendritic cell subset (Fig. 4c and Supplementary Fig. 5b)^{7,8}. The absence of iNOS⁺ dendritic cells in *iNOS*^{-/-}, *Myd88*^{-/-}, germ-free and *Tlr2*^{-/-} *Tlr4*^{-/-} *Tlr9*^{-/-} mice correlates well with the impaired IgA production in the sera and intestinal contents of these mice (Fig. 4d and Supplementary Fig. 5c). Furthermore, IgA production, B-cell expression of TGF-βRII, and the numbers of IgA⁺ B cells and IgA-producing plasma cells were

restored in $iNOS^{-/-}$ mice when transferred with a non-T/B-cell fraction of $iNOS^{+/+}$ lamina propria cells and $iNOS^{+/+}$ lamina propria dendritic cells (Fig. 4e, f and Supplementary Fig. 5d). In contrast, these factors (IgA production, B-cell expression of TGF- β RII, and the numbers of IgA $^{+}$ B cells and IgA-producing plasma cells) remained low in $iNOS^{-/-}$ mice when transferred with a dendritic-cell-depleted non-T/B-cell fraction of $iNOS^{+/+}$ lamina propria cells, a non-T/B-cell fraction of $iNOS^{-/-}$ lamina propria cells, $iNOS^{-/-}$ lamina propria dendritic cells and $iNOS^{+/+}$ spleen dendritic cells (Fig. 4e). Collectively, these results demonstrate that TNF- α /iNOS-producing dendritic cells are essential for IgA production *in vivo*.

Our results show that IgA production is selectively reduced in $iNOS^{-/-}$ mice. iNOS controls B-cell expression of TGF- β RII in T-cell-dependent IgA CSR, as well as dendritic-cell-derived APRIL and BAFF, and B-1-cell-derived IgA production in T-cell-independent IgA CSR. Consistent with these observations, mice lacking TGF- β RII in B cells show a selective reduction of IgA in the sera and external secretions^{19,20}, and the serum IgA level and IgA response to mucosal antigen exposure are reduced in *April* $^{-/-}$ mice²³ and *Taci* $^{-/-}$ mice²⁴. With respect to the molecular mechanism, NO regulates expression of TGF- β RII on B cells in a phosphatidylinositol-3 kinase/AKT-dependent manner, whereas it influences APRIL secretion by dendritic cells through NFAT- and NF- κ B-dependent

pathways. Supporting our results, NF- κ B is critical for BAFF induction²⁵. In addition, we also found that the naturally occurring TNF- α /iNOS-producing dendritic cells comprised a portion of retinal dehydrogenase (RALDH $^{+}$) dendritic cells in the lamina propria (Supplementary Fig. 6a–c), which imprint gut-homing specificity on T and B cells by inducing the integrin $\alpha 4 \beta 7$ (refs 26, 27); the number of $\alpha 4 \beta 7^{+}$ T and B cells was unaffected in TNF- α /iNOS-producing dendritic-cell-deficient $iNOS^{-/-}$ mice (Supplementary Fig. 6d, e). Accordingly, we propose that TNF- α /iNOS-producing dendritic cells directly govern IgA class-switching of naive B cells, whereas RALDH $^{+}$ dendritic cells imprint gut-homing specificity of the post-CSR IgA $^{+}$ B cells.

Impaired IgA production and intestinal epithelial cell homeostasis were observed in germ-free mice under steady-state conditions^{10,28}, and dendritic cells that reside between the intestinal epithelial cells sample commensal bacteria at the epithelial surface and participate in IgA induction in the MALT^{9,29,30}. Notably, we observed that iNOS is expressed in a subset of MALT dendritic cells, this expression is affected by toll-like-receptor-dependent recognition of commensal bacteria, and it has a critical role in IgA production. Our findings shed light on the important question of why IgA production, which is critical for the maintenance of gut homeostasis, takes place predominantly in the MALT.

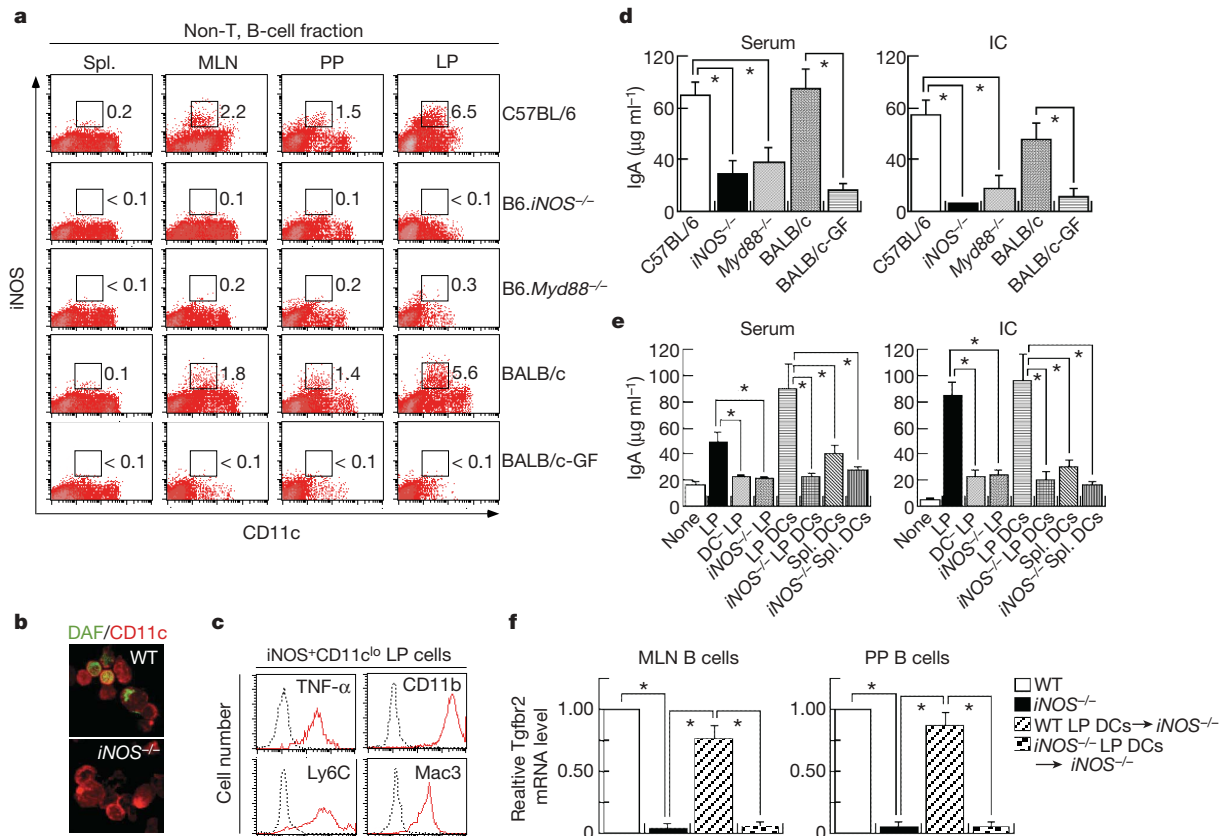


Figure 4 | Identification of iNOS $^{+}$ dendritic cells in the MALT.

a, CD3 ϵ^{-} B220 $^{-}$ cells from the indicated tissues of C57BL/6, B6.iNOS $^{-/-}$, B6.Myd88 $^{-/-}$, BALB/c and BALB/c germ-free mice were stained for CD11c and intracellular iNOS. **b**, NO detection in lamina propria dendritic cells by diaminofluorescein (DAF) staining. Original magnification, $\times 400$. **c**, Characterization of lamina propria iNOS $^{+}$ dendritic cells. Lamina propria iNOS $^{+}$ dendritic cells were stained for the indicated surface markers. Solid line, each marker staining; dashed line, isotype control staining. **d**, IgA level in the sera and intestinal content of C57BL/6 ($n = 5$), B6.iNOS $^{-/-}$ ($n = 5$), B6.Myd88 $^{-/-}$ ($n = 8$), BALB/c ($n = 5$) and BALB/c germ-free ($n = 12$) mice was measured by ELISA. **e**, IgA level in the sera and intestinal content of $iNOS^{-/-}$ mice transferred with a non-T/B-cell fraction of $iNOS^{+/+}$ lamina propria cells (LP), a dendritic-cell-depleted non-T/B-cell fraction of

$iNOS^{+/+}$ lamina propria cells (DC $^{-}$ LP), a non-T/B-cell fraction of $iNOS^{-/-}$ lamina propria cells ($iNOS^{-/-}$ LP), $iNOS^{+/+}$ lamina propria dendritic cells (LP DCs), $iNOS^{-/-}$ lamina propria dendritic cells ($iNOS^{-/-}$ LP DCs), $iNOS^{+/+}$ spleen dendritic cells (Spl. DCs), and $iNOS^{-/-}$ spleen dendritic cells ($iNOS^{-/-}$ Spl. DCs). Each group includes three mice. Two weeks after the transfer, the IgA level was determined by ELISA. **f**, TGF- β RII expression in naive B cells of C57BL/6, $iNOS^{-/-}$, $iNOS^{-/-}$ mice transferred with wild-type lamina propria dendritic cells (WT LP DCs) and $iNOS^{-/-}$ lamina propria dendritic cells ($iNOS^{-/-}$ LP DCs). Two weeks after the transfer, the TGF- β RII expression was evaluated by real-time PCR. Data are representative of three experiments. Asterisks in relevant panels indicate unpaired Student's *t*-test; *P* value is < 0.05 . Error bars represent s.d.

METHODS SUMMARY

Mice. C57BL/6 mice (B6 mice) and BALB/c mice were obtained from Clea. B6.iNOS^{-/-} mice were obtained from Taconic Farms, Inc. B6.Rag2^{-/-} mice were provided by Central Laboratories for Experimental Animals. F₂ ((C57BL/6 × iNOS^{-/-}) F₁ × iNOS^{-/-}) offspring were made by crossing C57BL/6 mice and iNOS^{-/-} mice. B6.iNOS^{-/-}Rag2^{-/-} mice were generated by crossing B6.iNOS^{-/-} mice and B6.Rag2^{-/-} mice. BALB/c germ-free mice were obtained from Yakult Central Institute for Microbiological Research. B6.Tlr2^{-/-}Tlr4^{-/-}Tlr9^{-/-} mice were made by crossing each Tlr^{-/-} mouse. All mice except for the germ-free mice were maintained in our SPF animal facility using the Micro-VENT system (Allentown Caging Equipment Company), and all animal experiments were done with the approval of the Institutional Animal Care Committee of the Akita University.

Reagents and *in vivo* treatment. To examine the requirement of iNOS and NO for IgA production, wild-type mice were given 2.5% (w/v) aminoguanidine (Calbiochem), 4.5 mM L-N⁶-(1-iminoethyl)-lysine (Wako), and 3 mg day⁻¹ carboxy-PTIO (2-(4-carboxyphenyl)-4,4,5,5-tetramethylimidazoline-1-oxyl-3-oxide) (Dojindo) intraperitoneally and in drinking water for 7–14 days.

Immunization. Mice were immunized intraperitoneally with 100 µg alum-precipitated ovalbumin (Sigma) (OVA/alum), boosted intraperitoneally with 50 µg OVA/alum 21 days later, and samples from the sera were collected 10 days after booster for estimation of anti-OVA antibodies.

Measurement of immunoglobulin levels. Samples were collected from sera, intestinal content, faecal pellets and culture supernatants. The levels of IgM, IgG, IgG1, IgG2b, IgG2c and IgG3 were determined using the mouse ELISA quantification kit (Bethyl).

PCR. For semi-quantitative RT-PCR analysis, serial dilutions of the cDNA templates were amplified using specific primers on a DNA thermal cycler (MJ Research). Primers for pIgR, AID, α-GT, TGF-βRI, TGF-βRII, iNOS and GAPDH were obtained from Hokkaido System Science Co. Real-time PCR analysis was performed using LightCycler (Roche) to measure SYBR green (Roche) incorporation. Primers for TGF-βRII, Smad3, Runx3, APRIL, BAFF and β-actin were obtained from Nihon gene research Laboratories Inc. Relative expression of mRNA was normalized to β-actin mRNA levels within each sample.

Statistical analysis. The statistical significance of the obtained values was evaluated by Student's *t*-test. A *P* value <0.05 was considered significant.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 May; accepted 19 June 2007.

- Brandtzaeg, P. *et al.* Regional specialization in the mucosal immune system: what happens in the microcompartments? *Immunol. Today* **20**, 141–151 (1999).
- van Egmond, M. *et al.* IgA and the IgA Fc receptor. *Trends Immunol.* **22**, 205–211 (2001).
- Mostov, K. E. Transepithelial transport of immunoglobulins. *Annu. Rev. Immunol.* **12**, 63–84 (1994).
- Matsunaga, T. & Rahman, A. What brought the adaptive immune system to vertebrates?—The jaw hypothesis and the seahorse. *Immunol. Rev.* **166**, 177–186 (1998).
- Fagarasan, S. & Honjo, T. Intestinal IgA synthesis: regulation of front-line body defences. *Nature Rev. Immunol.* **3**, 63–72 (2003).
- Suzuki, K. *et al.* Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. *Proc. Natl Acad. Sci. USA* **101**, 1981–1986 (2004).
- Serbina, N. V., Salazar-Mather, T. P., Biron, C. A., Kuziel, W. A. & Pamer, E. G. TNF/ iNOS-producing dendritic cells mediate innate immune defense against bacterial infection. *Immunity* **19**, 59–70 (2003).
- Serbina, N. V. & Pamer, E. G. Monocyte emigration from bone marrow during bacterial infection requires signals mediated by chemokine receptor CCR2. *Nature Immunol.* **7**, 311–317 (2006).

- Macpherson, A. J. & Uhr, T. Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria. *Science* **303**, 1662–1665 (2004).
- Macpherson, A. J. *et al.* A primitive T cell-independent mechanism of intestinal mucosal IgA responses to commensal bacteria. *Science* **288**, 2222–2226 (2000).
- Litinskiy, M. B. *et al.* DCs induce CD40-independent immunoglobulin class switching through BLyS and APRIL. *Nature Immunol.* **3**, 822–829 (2002).
- Castigli, E. *et al.* TACI and BAFF-R mediate isotype switching in B cells. *J. Exp. Med.* **201**, 35–39 (2005).
- Mestecky, J. & McGhee, J. R. Immunoglobulin A (IgA): molecular and cellular interactions involved in IgA biosynthesis and immune response. *Adv. Immunol.* **40**, 153–245 (1987).
- Fagarasan, S., Kinoshita, K., Muramatsu, M., Ikuta, K. & Honjo, T. *In situ* class switching and differentiation to IgA-producing cells in the gut lamina propria. *Nature* **413**, 639–643 (2001).
- Bogdan, C. Nitric oxide and the immune response. *Nature Immunol.* **2**, 907–916 (2001).
- Fujihashi, K. *et al.* γ/δ T cell-deficient mice have impaired mucosal immunoglobulin A responses. *J. Exp. Med.* **183**, 1929–1935 (1996).
- Shimada, S.-I. *et al.* Generation of polymeric immunoglobulin receptor-deficient mouse with marked reduction of secretory IgA. *J. Immunol.* **163**, 5367–5373 (1999).
- Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
- Cazac, B. B. & Roes, J. TGF-β receptor controls B cell responsiveness and induction of IgA *in vivo*. *Immunity* **13**, 443–451 (2000).
- Borsutzky, S., Cazac, B. B., Roes, J. & Guzman, C. A. TGF-β receptor signaling is critical for mucosal IgA responses. *J. Immunol.* **173**, 3305–3309 (2004).
- Park, S. R., Lee, J. H. & Kim, P. H. Smad3 and Smad4 mediate transforming growth factor-β1-induced IgA expression in murine B lymphocytes. *Eur. J. Immunol.* **31**, 1706–1715 (2001).
- Fainaru, O. *et al.* Runx3 regulates mouse TGF-β-mediated dendritic cell function and its absence results in airway inflammation. *EMBO J.* **23**, 969–979 (2004).
- Castigli, E. *et al.* Impaired IgA class switching in APRIL-deficient mice. *Proc. Natl Acad. Sci. USA* **101**, 3903–3908 (2004).
- von Bulow, G.-U., van Deursen, J. M. & Bram, R. J. Regulation of the T-independent humoral response by TACI. *Immunity* **14**, 573–582 (2001).
- He, B., Raab-Traub, N., Casali, P. & Cerutti, A. EBV-encoded latent membrane protein 1 cooperates with BAFF/BLyS and APRIL to induce T cell-independent Ig heavy chain class switching. *J. Immunol.* **171**, 5215–5224 (2003).
- Iwata, M. *et al.* Retinoic acid imprints gut-homing specificity on T cells. *Immunity* **21**, 527–538 (2004).
- Mora, J. R. *et al.* Generation of gut-homing IgA-secreting B cells by intestinal dendritic cells. *Science* **314**, 1157–1160 (2006).
- Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* **118**, 229–241 (2004).
- Rescigno, M. *et al.* Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nature Immunol.* **2**, 361–367 (2001).
- Niess, J. H. *et al.* CX3CR1-mediated dendritic cell access to the intestinal lumen and bacterial clearance. *Science* **307**, 254–258 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Onodera for animal care; K. Yamashita for experimental support; K. Honda, M. Muramatsu and M. Miyasaka for discussions; M. Nanno for BALB/c germ-free mice; T. Tsubata for B6.µMt mice; and K. Takatsu for monoclonal antibody H-7. This work was supported by Yakult Bio-Science Foundation (to T.O.), a Sasakawa Scientific Research Grant from The Japan Science Society (to H. Tezuka), a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports and Culture of Japan (T.O.), and a Grant-in-Aid for Young Scientists (B) (to H. Tezuka).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.O. (tohteki@med.akita-u.ac.jp).

METHODS

Mice. *B6.Tcrb^{-/-}Tcrd^{-/-31}*, *B6.Myd88^{-/-32}* and *B6.μMt³³* mice were as described.

Measurement of immunoglobulin levels. The level of IgA was measured by enzyme-linked immunosorbent assay (ELISA) as described previously³¹.

Cell preparation. To prepare naive B cells, spleen cells were labelled with PE-conjugated anti-mouse IgD mAb (eBioscience) and anti-PE MicroBeads (MiltenyiBiotec). IgD⁺ B cells were isolated using AutoMACS (MiltenyiBiotec) and used as naive B cells (>97% IgD⁺). Likewise, B220⁺ cells isolated from spleen, MLN, and PP were used as conventional B cells (>98% B220⁺). To prepare B-1 cells, peritoneal cells were labelled with PE-BM8 (anti-F4/80), PE-B3B4 (anti-CD23), and Thy1.2 (anti-CD90.2) MicroBeads and subsequently with anti-PE MicroBeads. The negative fraction was used as B-1 cells. To prepare BM DCs, BM cells were cultured in the presence of 10 ng ml⁻¹ recombinant mouse GM-CSF (R&D Systems) for 6 days by replacing medium every 3 days. After 6 days of culture, BM cells were further stimulated with 20 ng ml⁻¹ recombinant mouse TNF-α for 2 days (R&D Systems). The cells were then labelled with N418-MicroBeads (anti-mouse CD11c) (MiltenyiBiotec) and isolated using AutoMACS.

To prepare the intestinal intraepithelial lymphocytes (IELs), PPs, fatty tissues and mesentery were removed from the small intestine. The small intestine was opened longitudinally and cut into 5-mm pieces. The pieces were washed gently and shaken for 45 min at 37 °C in RPMI 1640 medium containing 5% FCS, 100 U ml⁻¹ penicillin G, and 100 μg ml⁻¹ streptomycin. Whole-intestinal-cell suspensions, which include IECs and IELs, were applied to discontinuous Percoll density gradients (Amersham Bioscience) of 25%, 40% and 75%, and cells at the interface between 40% and 75% were collected and used as IELs. After removal of IECs and IELs, the remaining tissue was further digested with collagenase (type IV, Sigma), and cells at the interface between 30% and 60% of Percoll gradient were collected and used as LP cells. To purify DCs, collagenase digested spleen, MLN, PP and LP cells were labelled with N418-MicroBeads (anti-mouse CD11c) (MiltenyiBiotec) and isolated using AutoMACS. These purified CD11c⁺ cells (>95% CD11c⁺ I-A⁺) were used as BM, spleen, MLN, PP and LP DCs.

Cell culture. To examine Ig secretion, B cells were cultured with anti-CD40 mAb (2 μg ml⁻¹; eBioscience), LPS (2 μg ml⁻¹, Sigma), TGF-β1 (1 ng ml⁻¹, RDI), mouse IL-5 (100 ng ml⁻¹, R&D Systems), mouse IL-15 (100 ng ml⁻¹, RDI), mouse APRIL (100 ng ml⁻¹, RDI), and anti-TGFβ1 mAb (10 μg ml⁻¹, R&D Systems), and the culture supernatants were collected for ELISA. A cell proliferation assay was also performed in some experiments, in which the cells were pulsed with 1 μCi per well ³H-thymidine (PerkinElmer Life Science) for the last 16 h of a 72-h culture, and ³H-thymidine incorporation was measured by a β-counter (Packard). In some experiments, 1 × 10⁶ T cells that had been activated with anti-CD3ε mAb (1 μg ml⁻¹) and anti-CD28 mAb (1 μg ml⁻¹) (eBioscience), were co-cultured with 5 × 10⁵ splenic naive B cells in the presence of TGF-β1 for 7 days. Likewise, 5 × 10⁵ B cells were co-cultured with 1.25–2.5 × 10⁵ DCs in the presence or absence of APRIL, TACI-Ig (3.3 μg ml⁻¹, R&D Systems) or BCMA-Ig (3.3 μg ml⁻¹, R&D Systems) for 7 days. Splenic IgD⁺ B cells were cultured with LY294002 (PI3 kinase inhibitor, Calbiochem) or Akt inhibitor (Calbiochem) for 1 h before addition of 100 μM spermine NONOate (Alexis Biochemicals) for 18 h. LP DCs were cultured with 5 μM BAY11-7082 (NF-κB inhibitor, Calbiochem), 5 μM 11R-VIVIT (NFAT inhibitor peptide, RRRRRRRRRR-GGG-MAGPVIVITGPHEE, Sigma Genosys)³⁴, 5 μM 11R-VEET (control peptide, RRRRRRRRRR-GGG-MAGPPHIVEETGPHEE, Sigma Genosys)³⁴, or 500 μM L-NIL for 1 h before the addition of 100 μM spermine NONOate or 1 μM CpG-A (5'-TCCATGACGTTCTCTGATGCT-3', Hokkaido System Science) for 18 h.

Antibodies and flow-cytometric analysis. Cells were stained with the following mAbs: fluorescein isothiocyanate (FITC)-N418 (anti-CD11c), phycoerythrin (PE)-RA3-B2 (anti-B220), PE-53-6.7 (anti-CD8α), PE-B3B4 (anti-CD23), PE-BM8 (anti-F4/80), biotin-UC7-13D5 (anti-TCRγδ), biotin-1C10 (anti-CD40), biotin-N418, biotin-M1/70 (anti-CD11b), biotin-16-10A1 (anti-CD80), biotin-GL1 (anti-CD86), biotin-RB6-8C5 (anti-Gr-1), biotin-53-6.7 (anti-CD8α), biotin-MB19-1 (anti-CD19), biotin-RA3-B2, biotin-53-2.1 (anti-Thy1.2), biotin-2E7 (anti-CD103), biotin-DX5 (anti-DX5), biotin-NLDC-145 (anti-CD205), and biotin-4B12 (anti-CCR7). These mAbs were purchased from eBioscience. FITC-C10-3 (anti-IgA), FITC-1A8 (anti-Ly6G), and biotin-AL-21 (anti-Ly6C) were purchased from BD Pharmingen. Biotin-H-7 (anti-IL-5Rα) was provided by T. Kouro and K. Takatsu of Tokyo University. Biotinylated mAbs were detected with streptavidin-PE-Cy5 (eBioscience). For intracellular iNOS and Mac-3 staining, CD3ε⁺B220⁺ cells of the spleen, MLN, PP and LP were stained with biotin N418 and streptavidin-PE-Cy5, and were fixed and

permeabilized with Cytofix/Cytoperm solution (BD Pharmingen). The cells were then stained intracellularly with PE-N20 (anti-iNOS, Santa Cruz) and FITC-M3/84 (anti-Mac-3) (eBioscience). For intracellular TNF-α staining, CD3ε⁺B220⁺ cells of the LP were cultured with 50 ng ml⁻¹ PMA (Wako Pure Chemical Industries), 500 ng ml⁻¹ ionomycin (Sigma), and GolgiPlug (BD Pharmingen) for 4 h, after which they were stained with biotin N418 and PE-streptavidin-Cy5 (eBioscience), and fixed and permeabilized as described above. The cells were then stained intracellularly with PE-N20 and FITC-MP6-XT22 (anti-TNF-α). FITC-rat IgG1 (eBioscience) was also included as isotype controls for Mac-3 and TNF-α. Samples were analysed on a FACSCalibur using the CELLQuest program (Becton Dickinson).

Immunocytochemical analysis. LP DCs and splenic DCs were fixed and permeabilized with Cytofix/Cytoperm solution (BD Pharmingen). These cells were stained with rabbit anti-mouse RALDH1 or RALDH2 (prepared in-house), and PE-N-20 (anti-iNOS). For RALDH staining, cells were further stained with FITC-conjugated anti-rabbit IgG (Jackson ImmunoResearch) and AlexaFluor 488-conjugated anti-FITC (Molecular Probes). To detect intracellular NO, LP DCs were incubated with 10 μM diaminofluorescein (DAF)-FM diacetate (Daichi Pure Chemicals) and then stained with PE-N418 (anti-CD11c). Stained cells were mounted on slides and analysed by fluorescence microscopy.

RT-PCR. Specific primer pairs used for RT-PCR were as follows: pIgR sense (5'-TTGTTACGCTCTTGGTAACGTG-3') and pIgR antisense (5'-ACAGGCC-TCGGTTACTGGTACC-3'); AID sense (5'-ATATGGACAGCCTTCTGATG-AAGC-3') and AID anti-sense (5'-TCAAAATCCCAACATACGAAATGC-3'); αGT sense (5'-CTACCATAGGGGAAGATAGCCT-3') and αGT anti-sense (5'-TAATCGTGAATCAGGCAG-3'); TGF-βRI sense (5'-ATCCATCACTAG-ATCGCCCT-3') and TGF-βRI anti-sense (5'-CGATGGATCAGAAGGTA-CAAGA-3'); TGF-βRII sense (5'-CGTGTGGAGGAAGAACAACA-3') and TGF-βRII anti-sense (5'-TCTCAAACCTGCTCTGAGGTG-3'); iNOS sense (5'-CGTTGGATTGGAGCAGAAAGTG-3') and iNOS anti-sense (5'-CATGC-AAAATCTCTCCACTGCC-3'); GAPDH sense (5'-ACCACAGTCCATGCC-ATCAC-3') and GAPDH anti-sense (5'-TCCACCACCCTGTTGCTGTA-3').

Specific primer pairs used for real-time PCR analysis were as follows: TGF-βRII sense (5'-CCTACTCTGTCTGTGGATGA-3') and TGF-βRII anti-sense (5'-GCTCGTAATCCTTCACTTCTC-3'); Smad3 sense (5'-GCACAGCCAC-CATGAATTAC-3') and Smad3 anti-sense (5'-GCACCAACTGGAGG-TAGA-3'); Runx3 sense (5'-ACCACGAGCCACTTCAGCAG-3') and Runx3 anti-sense (5'-CGATGGTGTGGCGCTGTA-3'); APRIL sense (5'-TCACAA-TGGGTCAGGTGGTATC-3') and APRIL anti-sense (5'-TGTAATGAAA-GACACCTGCATGT-3'); BAFF sense (5'-TGCTATGGGTGCATGTCACA-3') and BAFF anti-sense (5'-GGCAGTGTGTTGGGCATATTC-3'); β-actin sense (5'-GATGACGATATCGCTGCGCTG-3') and β-actin anti-sense (5'-GTACGACCAGAGGCATACAGG-3'). Relative expression of mRNA was normalized to β-actin mRNA levels within each sample.

Adoptive transfer. To examine the IgA productivity of *iNOS^{-/-}* B cells, naive B cells (a mixture of 2 × 10⁷ splenic B cells, 1.5 × 10⁶ MLN B cells, and 3 × 10⁶ PP B cells) of WT and *iNOS^{-/-}* mice were intravenously transferred into *μMt* hosts. To examine the function of *iNOS^{-/-}* T cells, 1 × 10⁷ spleen T cells of WT and *iNOS^{-/-}* mice were transferred into *Tcrb^{-/-}Tcrd^{-/-}* mice. To examine the role of *iNOS^{+/-}* DCs in IgA production, 2 × 10⁶ of a non-T/B fraction of *iNOS^{+/-}* LP cells, a DC-depleted non-T/B fraction of *iNOS^{+/-}* LP cells, a non-T/B fraction of *iNOS^{-/-}* LP cells, and 2 × 10⁶ *iNOS^{+/-}* LP DCs, *iNOS^{+/-}* SPL DCs were transferred into *iNOS^{-/-}* mice. Two to four weeks after the transfer, the IgA level was determined by ELISA. To examine the trafficking and survival of IgA⁺ B cells in *iNOS^{-/-}* hosts, PP IgA⁺ B cells (4 × 10⁶ cells) or LP IgA⁺ B cells (2 × 10⁶ cells) of B6.SJL (CD45.1⁺) mice were intravenously transferred into *iNOS^{+/-}* *Rag2^{-/-}* (WT), and *iNOS^{-/-}* *Rag2^{-/-}* (*iNOS^{-/-}*) hosts (CD45.2⁺). Sixteen hours (migration of the PP IgA⁺ B cells) and 6 days (survival of the LP IgA⁺ B cells) after the transfer, LP cells isolated from hosts were stained with FITC-104 (anti-CD45.2) and PE-A20 (anti-CD45.1) and analysed on FACSCalibur using the CELLQuest program (Becton Dickinson).

- Nishiyama, Y. *et al.* Homeostatic regulation of intestinal villous epithelia by B lymphocytes. *J. Immunol.* **168**, 2626–2633 (2002).
- Adachi, O. *et al.* Targeted disruption of the *MyD88* gene results in loss of IL-1- and IL-18-mediated function. *Immunity* **9**, 143–150 (1998).
- Kitamura, D., Roes, J., Kuhn, R. & Rajewsky, K. A B cell-deficient mouse by targeted disruption of the membrane exon of the immunoglobulin μ chain gene. *Nature* **350**, 423–426 (1991).
- Noguchi, H., Matsushita, M., Okitsu, T. & Matsui, H. A new cell-permeable peptide allows successful allogeneic islet transplantation in mice. *Nature Med.* **10**, 305–309 (2004).

LETTERS

An IRF8-binding promoter variant and AIRE control *CHRNA1* promiscuous expression in thymus

Matthieu Giraud^{1,2}, Richard Taubert³, Claire Vandiedonck¹, Xiayi Ke⁴, Matthieu Lévi-Strauss¹, Franco Pagani⁵, Francisco E. Baralle⁵, Bruno Eymard⁶, Christine Tranchant⁷, Philippe Gajdos⁸, Angela Vincent², Nick Willcox², David Beeson², Bruno Kyewski³ & Henri-Jean Garchon¹

Promiscuous expression of tissue-restricted auto-antigens in the thymus imposes T-cell tolerance and provides protection from autoimmune diseases^{1–3}. Promiscuous expression of a set of self-antigens occurs in medullary thymic epithelial cells^{4,5} and is partly controlled by the autoimmune regulator (AIRE), a nuclear protein for which loss-of-function mutations cause the type 1 autoimmune polyendocrine syndrome^{6,7}. However, additional factors must be involved in the regulation of this promiscuous expression. Here we describe a mechanism controlling thymic transcription of a prototypic tissue-restricted human auto-antigen gene, *CHRNA1*. This gene encodes the α -subunit of the muscle acetylcholine receptor, which is the main target of pathogenic auto-antibodies in autoimmune myasthenia gravis^{8,9}. On re-sequencing the *CHRNA1* gene, we identified a functional bi-allelic variant in the promoter that is associated with early onset of disease in two independent human populations (France and United Kingdom). We show that this variant prevents binding of interferon regulatory factor 8 (IRF8) and abrogates *CHRNA1* promoter activity in thymic epithelial cells *in vitro*. Notably, both the *CHRNA1* promoter variant and AIRE modulate *CHRNA1* messenger RNA levels in human medullary thymic epithelial cells *ex vivo* and also in a transactivation assay. These findings reveal a critical function of AIRE and the interferon signalling pathway in regulating quantitative expression of this auto-antigen in the thymus, suggesting that together they set the threshold for self-tolerance versus autoimmunity.

To understand better the relationship between genetic variants, thymic auto-antigen expression and autoimmune susceptibility we

undertook a comprehensive study of the polymorphism of the *CHRNA1* locus. Through re-sequencing in 16 individuals, we identified 70 variants that were then genotyped in 45 patients. We selected 11 tagging single nucleotide polymorphisms (SNPs), 5 of which represented allelic groups (A–E) (Supplementary Figs 1 and 2, and Supplementary Table 1). In 180 French myasthenia gravis patients overall, the allele and genotype frequencies were similar to those in 151 population-matched controls (Supplementary Table 2), but the age of onset of myasthenia gravis was influenced significantly by the B SNP group (as shown by multiple linear regression, $P = 0.007$). Exclusion of the upper and lower 5% of onset-age distribution did not alter the findings ($P = 0.00046$), ruling out a bias by extremes.

Group B included 12 tightly associated SNPs, three of which (rs2600683, rs16862847 and rs2600686) were potentially functional but genetically indistinguishable (Table 1 and Supplementary Table 3). The rs16862847 variant, located 478 nucleotides upstream of the translation initiation codon, has the potential to influence acetylcholine receptor (AChR) expression. Its minor G allele disrupts a unique interferon-stimulated response element that is a consensus binding motif for IRF8 (similarity score of 0.94 for a threshold of 0.81) (ref. 10). Previously known as interferon consensus sequence binding protein (ICSBP), this transcription factor is involved in signalling of interferons and also modulates their expression^{11,12}. Genotyping of rs16862847 in the French data set showed an increased minor allele frequency in myasthenia gravis patients with an age at disease onset in the lower quartile (odds ratio, 2.35; 95% confidence interval (CI), 1.31–4.23; $P = 0.0049$) compared to the control group,

Table 1 | Association of rs16862847 G with early onset of disease in French and UK myasthenia gravis patients

Population	Group	N	G allele			Genotype			AG + GG/AA	
			Frequency	Odds ratio (95% CI)	P-value	AA	AG	GG	Odds ratio (95% CI)	P-value
France	P _{LQ}	50	0.23	2.35 (1.31–4.23)	0.0049	27	23	0	3.3 (1.67–6.52)	0.0008
	P _{UTQ}	130	0.11			104	24	2		
	C	151	0.11			120	28	3		
United Kingdom	P _{LQ}	46	0.21	2.01 (1.04–3.87)	0.029	29	15	2	2.08 (0.98–4.4)	0.044
	P _{UTQ}	104	0.11			87	12	5		
	C	109	0.11			85	23	1		
Combined	P _{LQ}	96	0.22	2.19 (1.41–3.39)	0.00048	56	38	2	2.66 (1.6–4.41)	0.00015
	P _{UTQ}	234	0.11			191	36	7		
	C	260	0.11			205	51	4		

P_{LQ} indicates patients with early onset of myasthenia gravis, defined by age at the onset of disease in the lower quartile of the distribution in each cohort (≤ 21 yr in French patients, including 8 patients of age 21, and ≤ 18 yr in UK patients, including 10 patients of age 18). P_{UTQ} indicates patients in the upper three quartiles of the distribution. C indicates matched controls. Odds ratios were determined relative to the control populations. The exact P -values were tested two-sided in the French cohort, and one-sided in the replication UK cohort. The combined data set was tested two-sided using stratified 2×2 tables. The data also fitted a partially (or semi-) dominant model (Cochran–Armitage test with strata: $P = 4 \times 10^{-4}$; odds ratio for AG versus AA genotypes, 2.27, 95% CI, 1.4–3.69; for GG versus AA genotypes, 5.15, 95% CI, 1.96–13.6).

¹Inserm, U580, 75015 Paris, France; Université Paris-Descartes, 75015 Paris, France. ²Neuroscience Group, Weatherall Institute of Molecular Medicine, Oxford University, Headington, Oxford OX3 9DS, UK. ³Division of Developmental Immunology, Tumor Immunology Program, German Cancer Research Centre, INF 280, D-69120 Heidelberg, Germany. ⁴Centre for Integrated Genomic Medical Research (CIGMR) and Arthritis Research Campaign (ARC) Epidemiology Unit, University of Manchester, Manchester M13 9PT, UK. ⁵International Centre for Genetic Engineering and Biotechnology Padriciano 99, 34012 Trieste, Italy. ⁶Institut de Myologie, Hôpital de la Salpêtrière, 75013 Paris, France. ⁷Service de Neurologie, Hôpital Civil, CHRU, 67091 Strasbourg, France. ⁸Service de Réanimation, Hôpital Raymond Poincaré, 92380 Garches, France.

confirming the relationship to age at onset (Table 1). This association was replicated in a second series of 150 myasthenia gravis patients and 109 controls from the United Kingdom (t -test, $P = 0.0058$ and Table 1). Combining both data sets increased the evidence for an association of rs16862847 with an early onset of the disease ($P = 1.5 \times 10^{-4}$; common odds ratio, 2.66; 95% CI, 1.6–4.41).

To investigate the functional significance of rs16862847, we performed an electrophoretic mobility shift assay (EMSA) on radiolabelled oligonucleotides with nuclear extracts from the murine RAW264.7 macrophage cell line; this represents the myelo-monocytic lineage where expression of IRF8 was first described¹³. We found an allele-dependent molecular complex characterized by one or two bands (first two lanes, Fig. 1a). The upper band was specific for the major A allele whereas the lower one was shared by both rs16862847 alleles. Competition with unlabelled oligonucleotide or an irrelevant probe (nuclear factor of activated T-cells, NFAT) confirmed the specificity. Moreover, the A-allele-specific band bound IRF8, as it was supershifted by an anti-IRF8 antibody but not by an irrelevant anti-IRF1 antibody (third lane, Fig. 1b).

Among the other two candidate SNPs of group B, rs2600686 was situated in the third intron of *CHRNA1* and potentially created an intronic splicing processing element¹⁴. However, we found no evidence for a functional role of this variant (Supplementary Fig. 3). For the third potentially functional variant, rs2600683, the minor C allele created a DNA-binding motif for NFAT located 1,353 base pairs (bp) upstream of the *CHRNA1* translation initiation site. This motif was redundant but nonetheless an EMSA showed allele-specific binding of nuclear extracts (Supplementary Fig. 4).

To discriminate between the effects of rs2600683 and rs16862847 alleles on *CHRNA1* promoter activity, we tested the four possible allele combinations within the promoter (Fig. 2a) by transfection of reporter gene constructs into RAW264.7 cells and also a thymic epithelial cell (TEC) line derived from human neonatal thymic epithelium¹⁵. Both cell lines expressed IRF8, as verified by polymerase chain reaction with reverse transcription (RT-PCR) and western blotting (data not shown), and yielded similar findings. Figure 2 shows the results with the TEC line. Only rs16862847 allele variation influenced expression of the reporter luciferase gene; the minor G allele yielded a fivefold lower relative luciferase activity compared with the major A allele ($P = 1.5 \times 10^{-4}$ for the comparison of P1 versus P3 promoter activities with the Mann–Whitney U -test;

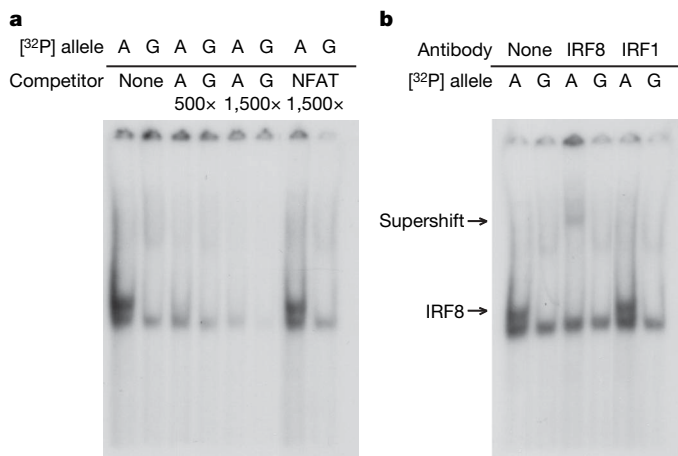


Figure 1 | rs16862847 allele variation affects IRF8 binding. **a**, EMSA. Nuclear extracts prepared from RAW264.7 macrophage cells were incubated with ³²P-labelled oligonucleotide probes corresponding to either allele of rs16862847, in the presence or absence of an excess of competing unlabelled oligonucleotide. An NFAT probe was used as an irrelevant competitor. The disease-accelerating G allele lacks the upper band of the complex. **b**, Supershift assay. Addition of an antibody against IRF8 but not against IRF1 clears the A-allele-specific band and yields a supershifted band.

Fig. 2b), similar to the background level of the promoterless construct. Notably, this allele-specific expression depended on IRF8, as shown by RNA interference (Fig. 2c); co-transfection with the P1 construct and short interfering (si)RNAs targeting *IRF8* decreased luciferase expression to the level of an anti-luciferase siRNA pool. Moreover, treatment of the cells with interferon (IFN)- γ upregulated the luciferase activity by a factor of 1.57 ($P = 6 \times 10^{-4}$), but only in the presence of the major allele of rs16862847 (Fig. 2d). We verified the mRNA expression from the endogenous *CHRNA1* gene in the TEC line that was homozygous for the major A allele of rs16862847 (data not shown); using chromatin immunoprecipitation, we showed specific binding of IRF8 to the promoter at the *CHRNA1* locus ($P < 0.05$), but not to an irrelevant promoter (*GAPDH*) (Fig. 2e). Thus, in TECs, IRF8 binds to the *CHRNA1* promoter carrying the rs16862847 A allele and enhances its transcriptional activity.

Next, we investigated the expression of *CHRNA1* *ex vivo* in relation to rs16862847. We also tested for an influence of *AIRE*. Messenger RNA levels of both genes were determined in samples of medullary (m)TECs freshly purified from the thymi of 25 (non-autoimmune) infants undergoing thoracic surgery. The levels varied among individuals but overall were closely correlated (Spearman's $R = 0.71$,

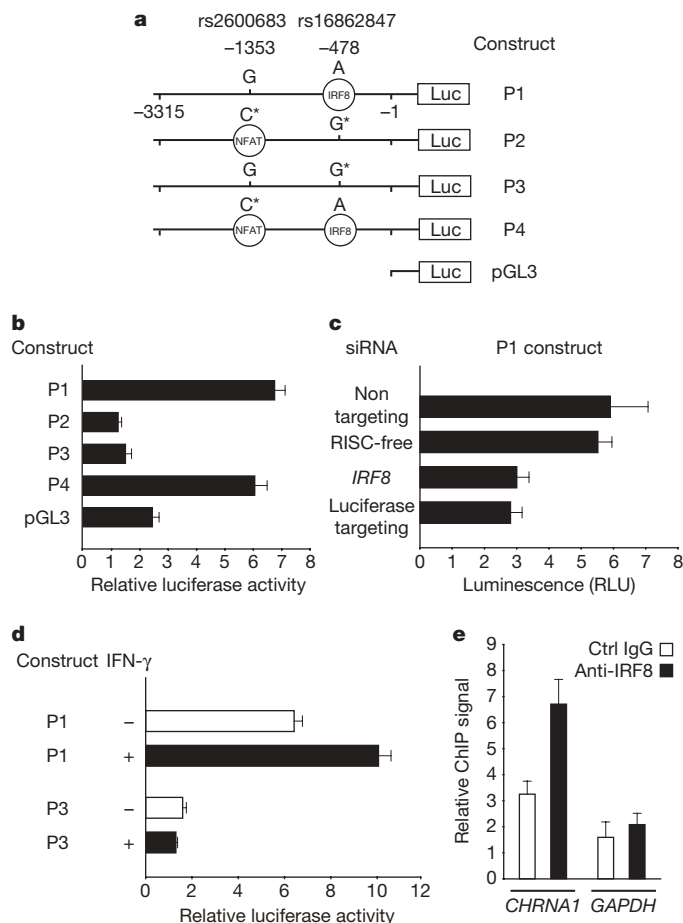


Figure 2 | Characterization of IRF8-dependent *CHRNA1* transcriptional activity in the TEC line. **a**, Disease-associated alleles (asterisks) at rs2600683 and rs16862847 sites were introduced by site-directed mutagenesis of P1, the common wild-type promoter, to generate P2–P4 constructs. **b**, Assay of promoter activities by relative luciferase activity. **c**, Effect of siRNA knockdown of *IRF8* on P1 promoter activity. **d**, Specific effect of IFN- γ on P1 construct expression. For **b–d**, data represent mean \pm s.e.m. of eight transfections for each plasmid or condition. **e**, Binding of IRF8 to the endogenous *CHRNA1* promoter detected by ChIP. The *GAPDH* promoter was analysed as a negative control for IRF8. Data represent mean \pm s.e.m. of quadruplicate real-time PCR.

$P = 7 \times 10^{-5}$; Fig. 3a, black line), suggesting a role for AIRE in controlling *CHRNA1* transcription. A similar relationship was found in murine mTECs (Supplementary Fig. 5). The correlation in human mTECs was preserved in subjects homozygous for the rs16862847 A allele (Spearman's $R = 0.87$, $P = 1 \times 10^{-5}$) but was not significant in those with one or two G alleles. Moreover, G-allele-carrying individuals showed a 1.8-fold lower mean level of *CHRNA1* mRNA compared with AA homozygotes (Fig. 3b; $P = 0.076$ for Student's *t*-test and $P = 0.0062$ when adjusted for AIRE level using a linear model). Notably, *IRF8* mRNA was expressed in both cortical and medullary TECs, displayed little inter-individual variation, and did not correlate with AIRE or *CHRNA1* mRNA levels (Supplementary Fig. 6).

To confirm the relationship between AIRE and *CHRNA1* mRNA expression and the involvement of rs16862847, we conducted co-transfection experiments in the TEC line using an AIRE expression plasmid and the *CHRNA1* promoter luciferase reporter constructs. We found a strong influence of AIRE on transcription from the *CHRNA1* promoter for both rs16862847 alleles (Fig. 3c). Luciferase expression first increased almost linearly as a function of the amount of the AIRE expression plasmid transfected and then reached a plateau ($P < 1 \times 10^{-15}$), in contrast to the continuous increase of the AIRE message (Supplementary Fig. 7). Nevertheless, there was also a highly significant effect of rs16862847 variation ($P < 1 \times 10^{-10}$), as reflected by a twofold decrease of luciferase expression with the G allele compared to A at all doses of AIRE plasmid. Moreover, treatment with siRNA targeting *IRF8* significantly decreased the response of the promoter with the A allele but not with the G allele, indicating that *IRF8* accounts for a large part of the allelic difference in AIRE-driven *CHRNA1* promoter activity (Fig. 3d).

The molecular mechanisms controlling the promiscuous expression of tissue-restricted auto-antigens in TECs are presently ill-defined and expression seems to be regulated at several levels⁵. After the

discovery of AIRE, our results outline a molecular framework for the AIRE-dependent transcriptional control of such an auto-antigen in TECs. Several mechanisms of action have previously been considered for AIRE. First, direct binding of AIRE to *cis*-acting DNA motifs; second, formation of higher order transcriptional complexes (without DNA binding); and third, modification of chromatin structure. Recent data (for discussion, see ref. 16) and our present findings are most consistent with the second mode. AIRE together with IRF8 might constitute part of a higher order transcriptional complex; by decreasing the binding of such a complex to the *CHRNA1* promoter, the rs16862847 minor allele would also affect the co-transcriptional activity of AIRE and weaken the correlation between AIRE and *CHRNA1* expression levels in mTECs (Fig. 3a, c). This does not exclude an additional effect of AIRE that is independent of IRF8, as increasing doses of AIRE also upregulate *CHRNA1* transcription in the minor allele *in vitro* (Fig. 3c, d). Alternatively, the role of AIRE could be more indirect. On the basis of the recent finding that AIRE in cooperation with CREB-binding protein is able to activate transcription of IFN- β ^{17,18}, it is conceivable that a member of the interferon family under the control of AIRE might induce IRF8 expression (see Supplementary Discussion).

Although the difference in mean thymic *CHRNA1* mRNA levels between individuals carrying the A and the G alleles was less than twofold, similar differences between diabetes-resistant and diabetes-susceptible insulin variable number tandem repeat (VNTR) alleles are important^{19–22}. Such sensitivity of central tolerance to relatively modest variations of antigen dose has also been described in animal models of autoimmunity^{23–25} and has been explained by the threshold model of tolerance² (see Supplementary Discussion). Thus, our data demonstrate dual control of thymic expression of an auto-antigen and highlight the role of central tolerance in protection from organ-specific autoimmunity in humans. How the genetic variant influences the onset of myasthenia gravis requires further investigation.

METHODS SUMMARY

Patients and controls. A total of 330 unrelated Caucasian patients from two independent cohorts (180 French and 150 UK patients) were included in the study with their written informed consent and in keeping with the tenets of the Declaration of Helsinki. They fulfilled clinical and electro-myographical criteria for the diagnosis of acquired generalized myasthenia gravis. They had confirmed positive titres of anti-AChR auto-antibodies (>0.6 nM), determined by radio-immunoprecipitation. Only non-thymoma patients were included. The control groups matched both populations and included 151 French and 109 UK volunteering Caucasian blood donors. They were not matched for gender; however, gender did not influence the genotype distributions (Supplementary Table 4).

Molecular methods and data analysis. Genetic analysis of the *CHRNA1* locus, EMSA protocols, cloning of the *CHRNA1* promoter and reporter gene assays, RNAi experiments, chromatin immunoprecipitation (ChIP), preparation and purification of medullary thymic epithelial cells, quantitative RT-PCR, and statistical methods are described in detail in the Methods. For transfections and ChIP, the data shown are representative of at least three experiments with similar results. For ChIP, independent chromatin preparations were used in each experiment.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 December 2006; accepted 2 July 2007.

Published online 8 August 2007.

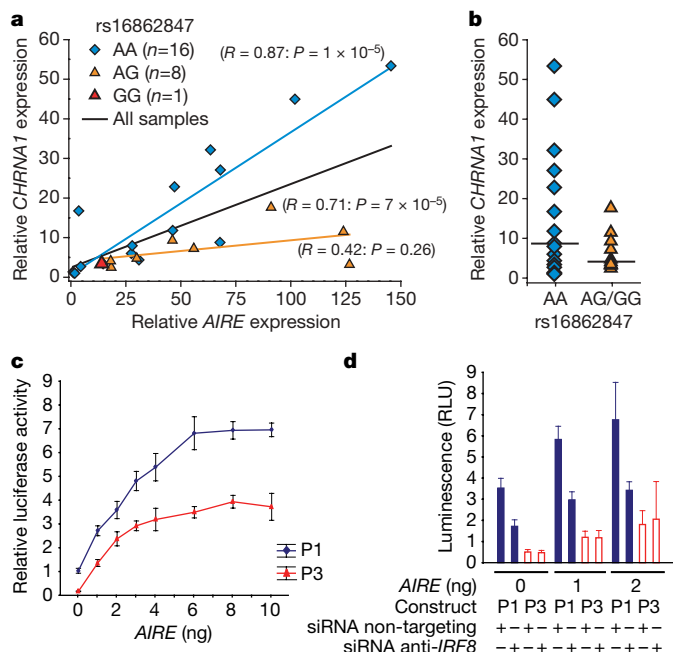


Figure 3 | Effect of rs16862847 and AIRE on *CHRNA1* expression levels in mTECs. **a**, Correlation between *CHRNA1* and AIRE message levels in all purified human mTEC samples (black line), in rs16862847 AA homozygotes (blue line, diamonds) or in rs16862847 AG or rs16862847 GG samples (orange line, triangles). **b**, Decreased *CHRNA1* expression associated with rs16862847 G allele. Horizontal bars show medians of relative expression, 8.37 for AA and 4.75 for AG or GG ($P = 0.0062$, adjusted for AIRE level). **c**, *In vitro* assay of the transactivation property of AIRE on *CHRNA1* P1 and P3 promoter constructs. **d**, RNAi knockdown of *IRF8* affects AIRE-stimulated activity of the *CHRNA1* promoter with rs16862847 A allele. For **c** and **d**, data represent mean \pm s.e.m. of four transfections for each condition.

- Mathis, D. & Benoist, C. Back to central tolerance. *Immunity* 20, 509–516 (2004).
- Liston, A., Lesage, S., Gray, D. H., Boyd, R. L. & Goodnow, C. C. Genetic lesions in T-cell tolerance and thresholds for autoimmunity. *Immunol. Rev.* 204, 87–101 (2005).
- Kyewski, B. & Klein, L. A central role for central tolerance. *Annu. Rev. Immunol.* 24, 571–606 (2006).
- Anderson, M. S. et al. Projection of an immunological self shadow within the thymus by the Aire protein. *Science* 298, 1395–1401 (2002).
- Derbinski, J. et al. Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. *J. Exp. Med.* 202, 33–45 (2005).
- Nagamine, K. et al. Positional cloning of the *APECED* gene. *Nature Genet.* 17, 393–398 (1997).

7. The Finnish-German APECED Consortium. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nature Genet.* **17**, 399–403 (1997).
8. Drachman, D. B. Myasthenia gravis. *N. Engl. J. Med.* **330**, 1797–1810 (1994).
9. Vincent, A. Unravelling the pathogenesis of myasthenia gravis. *Nature Rev. Immunol.* **2**, 797–804 (2002).
10. Tamura, T., Thotakura, P., Tanaka, T. S., Ko, M. S. & Ozato, K. Identification of target genes and a unique cis element regulated by IRF-8 in developing macrophages. *Blood* **106**, 1938–1947 (2005).
11. Barnes, B., Lubyova, B. & Pitha, P. M. On the role of IRF in host defense. *J. Interferon Cytokine Res.* **22**, 59–71 (2002).
12. Lohoff, M. & Mak, T. W. Roles of interferon-regulatory factors in T-helper-cell differentiation. *Nature Rev. Immunol.* **5**, 125–135 (2005).
13. Holtschke, T. *et al.* Immunodeficiency and chronic myelogenous leukemia-like syndrome in mice with a targeted mutation of the *ICSBP* gene. *Cell* **87**, 307–317 (1996).
14. Pagani, F. *et al.* A new type of mutation causes a splicing defect in *ATM*. *Nature Genet.* **30**, 426–429 (2002).
15. Fernandez, E. *et al.* Establishment and characterization of cloned human thymic epithelial cell lines. Analysis of adhesion molecule expression and cytokine production. *Blood* **83**, 3245–3254 (1994).
16. Kyewski, B. & Derbinski, J. Self-representation in the thymus: an extended view. *Nature Rev. Immunol.* **4**, 688–698 (2004).
17. Sillanpaa, N. *et al.* Autoimmune regulator induced changes in the gene expression profile of human monocyte-dendritic cell-lineage. *Mol. Immunol.* **41**, 1185–1198 (2004).
18. Pitkanen, J. *et al.* Cooperative activation of transcription by autoimmune regulator AIRE and CBP. *Biochem. Biophys. Res. Commun.* **333**, 944–953 (2005).
19. Vafiadis, P. *et al.* Insulin expression in human thymus is modulated by *INS* VNTR alleles at the *IDDM2* locus. *Nature Genet.* **15**, 289–292 (1997).
20. Pugliese, A. *et al.* The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the *INS* VNTR-*IDDM2* susceptibility locus for type 1 diabetes. *Nature Genet.* **15**, 293–297 (1997).
21. Sabater, L. *et al.* Insulin alleles and autoimmune regulator (AIRE) gene expression both influence insulin expression in the thymus. *J. Autoimmun.* **25**, 312–318 (2005).
22. Taubert, R., Schwendemann, J. & Kyewski, B. Highly variable expression of tissue-restricted self-antigens in human thymus: implications for self-tolerance and autoimmunity. *Eur. J. Immunol.* **37**, 838–848 (2007).
23. Chentoufi, A. A. & Polychronakos, C. Insulin expression levels in the thymus modulate insulin-specific autoreactive T-cell tolerance: the mechanism by which the *IDDM2* locus may predispose to diabetes. *Diabetes* **51**, 1383–1390 (2002).
24. Thebault-Baumont, K. *et al.* Acceleration of type 1 diabetes mellitus in proinsulin 2-deficient NOD mice. *J. Clin. Invest.* **111**, 851–857 (2003).
25. Liston, A. *et al.* Gene dosage-limiting role of Aire in thymic expression, clonal deletion, and organ-specific autoimmunity. *J. Exp. Med.* **200**, 1015–1026 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was funded by INSERM (Institut National de la Santé et de la Recherche Médicale) and the Association Française contre les Myopathies. M.G. was supported by a fellowship of the Fondation pour la Recherche Médicale in Oxford and of the European Molecular Biology Organization in Trieste. R.T. and B.K. have been supported by the DKFZ, the Sonderforschungsbereich 405 and EU-Thymaide; F.E.B. and F.P. by the Associazione Italiana Ricerca Cancro. D.B., N.W. and A.V. thank the UK MRC, Muscular Dystrophy Campaign and the Myasthenia Gravis Association for support. We thank the participating patients. We thank J. Newsom-Davis for clinical samples and expertise; S. Krumeich and G. Picarda for technical assistance; W. Savino for providing us with an aliquot of the TEC line; A. Harris for advice on EMSA; M. Lathrop for advice on genetic data analysis; P. Peterson for the gift of the AIRE expression plasmid; S. Hagl and members of the Department of Cardiac Surgery for providing human thymic tissue; K. Hexel and M. Scheuermann for cell sorting; G. Moldenhauer for providing monoclonal antibodies; and A. Kopp-Schneider for statistical analysis.

Author Contributions M.G. did the genomic experiments, EMSAs, made the reporter constructs, carried out transfections, RNAi and ChIP, analysed the data, and wrote the manuscript; R.T. carried out the experiments related to thymic gene expression; C.V. did genotyping and edited the manuscript; X.K. analysed the genetic data and edited the manuscript; M.L.-S. provided expertise to M.G. on gene cloning and proofread the manuscript; F.P. and F.E.B. provided expertise on RNA splicing analysis to M.G., when he was in Trieste, and proofread the manuscript; B.E., C.T. and P.G. recruited the French patients; A.V. and N.W. helped plan the project, provided access to the UK cohort and edited the manuscript; D.B. helped devise the experiments related to EMSAs and luciferase reporter constructs, supervised M.G. when he was in Oxford, and edited the manuscript; B.K. devised and supervised experiments related to promiscuous gene expression and wrote the manuscript; H.-J.G. coordinated the project, analysed the data and wrote the manuscript.

Author Information Newly identified SNPs (listed with an asterisk in Supplementary Table 1; accession numbers also listed) have been deposited in dbSNP. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.-J.G. (garchon@necker.fr).

METHODS

Polymorphism detection, genotyping and SNP tagging. We identified polymorphisms in *CHRNA1* by direct re-sequencing of genomic DNA from 13 French myasthenia gravis patients and 3 controls. This sample of 32 chromosomes provided a power of 80.1% to detect variants with a minor allele frequency of 0.05 at the 5% level of significance. Primers were selected with Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) to define 15 overlapping amplicons with a length up to 3 kb. These amplicons encompassed a genomic segment of 19.9 kb, including 3.3 kb upstream and 0.3 kb downstream of the *CHRNA1* gene (Supplementary Fig. 1a). Each amplicon was sequenced using both amplification and internal primers. Reaction products were migrated on an ABI 3100 automated sequencer (Applied Biosystems). We analysed sequencing chromatograms with the Phred/Phrap/Consed suite²⁶. Polymorphisms were detected with Polyphred and verified manually. SNPs were genotyped by multiplexing single-base chain extension reactions, using the SNaPshot system (Applied Biosystems). Insertions/deletions were genotyped by direct re-sequencing. This first step led us to identify 70 polymorphisms, 31 of which were new (Supplementary Table 1).

We then genotyped these 70 variants in 32 additional patients (45 patients in total). Of the 70 variants, 43 had a minor allele frequency >0.05. Tagging SNPs were selected using ldSelect²⁷ with an r^2 threshold at 0.75 and minor allele frequency threshold at 5%. Eleven tagSNPs were selected to represent five linkage disequilibrium (LD) groups, denoted A to E, and six independent SNPs (identified with a hash in Supplementary Table 1 and in Supplementary Fig. 2).

Finally, we genotyped the 11 tagSNPs in 180 French myasthenia gravis patients and 151 population-matched controls. Their genotypic frequencies did not deviate from Hardy–Weinberg proportions in either patients or controls. In the UK cohort, rs16862847 was genotyped by PCR-RFLP, through a *Bst*XI digestion of a 126-bp fragment amplified with a forward primer 5'-TCAGCCTAACACTGAAGCCAAGGAAATG-3' and a reverse primer: 5'-TGTGGGTGACTTGATTAGGG-3'.

Prediction of DNA-binding motifs. We used the MatInspector program from Genomatix (<http://www.genomatix.de>) to predict and compare DNA-binding motifs in promoter sequences with alternative alleles, and ModelInspector to list promoters including an interferon-stimulated response element. In intronic sequences we looked for regulatory elements in the literature, at splice site prediction of the Berkeley *Drosophila* Genome Project (http://www.fruitfly.org/seq_tools/splice.html) and at the alternative splicing workbench of the European Bioinformatics Institute (<http://www.ebi.ac.uk/asd-srv/wb.cgi>).

EMSA. We carried out preparation of nuclear extract from RAW264.7 cells and EMSA using the NucBuster Protein Extraction Kit and the EMSA Accessory Kit (Novagen). rs16862847 and rs2600683 probes were made by annealing of overlapping complementary DNA oligo-nucleotides. rs16862847 G allele, forward 5'-GTGAAGCAGAGGAAATGGAAGCCATGT-3', reverse 5'-GGAAACATGGCTTCAGTTCCATTTCCTCTGCT-3'; rs16862847 A allele, forward 5'-GTGAAGCAGAGGAAATGGAAGCCATGT-3', reverse 5'-GGAAACATGGCTTCAGTTTCATTCTCTGCT-3'; rs2600683 C allele, forward 5'-GTGAGAACATTTTCCACTCATGCCACT-3', reverse 5'-GGCCAGTGGCATGAGTGGAAGAAAATGTTTC-3'; rs2600683 G allele, forward 5'-GTGAGAACATTTTCCACTCATGCCACT-3', reverse 5'-GGCCAGTGGCATGAGTGGAAGAAAATGTTTC-3'. Underlined nucleotides correspond to the rs2600683 and rs16862847 alleles and italic sequences to the IRF8- and NFAT-binding motifs. We labelled probe overhangs with [α -³²P]dCTP. We incubated RAW264.7 nuclear extracts with 10 fmol of labelled probe for 30 min at 4 °C. For competition experiments, before adding to the extract, we mixed the labelled probe with a 500-fold or 1,500-fold molar excess of unlabelled probe or alternatively the NFAT probe 5'-CGCCCAAGAGGAAATTTGTTTCATA-3' (Santa Cruz Biotechnology). For supershift assays, we added 2 μ g of antibodies against IRF8 (C-19) or IRF1 (M-20) (Santa Cruz Biotechnology) for 20 min after incubation with the labelled probe. We separated protein–DNA complexes on a non-denaturing 6% Novex retardation gel (Invitrogen) in 0.5× TBE buffer. The gel was dried and auto-radiographed.

Promoter cloning and reporter gene assays. We amplified a fragment corresponding to nucleotides –3315 to –1 (relative to the ATG start codon +1) by PCR using the PfuUltra Hotstart DNA polymerase (Stratagene) of one individual (ID05 on Supplementary Fig. 2). This individual was homozygous for the upstream region common haplotype (44%) characterized by common alleles at each polymorphism position, including G for rs2600683 and A for rs16862847. We used primers tailed with *Nhe*I and *Xho*I restriction sites to sub-clone the fragment into the *Nhe*I and *Xho*I sites of the pGL3-Basic vector (Promega). Three vectors were derived by site-directed mutagenesis with the QuickChange II XL site-directed mutagenesis kit (Stratagene), inserting the minor allele, C for rs2600683, G for rs16862847 or both. We verified the four constructs by DNA sequencing. Maxipreps of each were made using the

EndoFree plasmid maxi kit (Qiagen). Transient transfection assays were conducted in a human thymic epithelial cell (TEC) line (clone P1.4D6)¹⁵ and in a murine macrophage cell line (RAW264.7) using the Lipofectamine 2000 transfection reagent (Invitrogen), in 96-well plates with 4–8 replicate wells per condition per experiment in at least three independent experiments. Transfections were carried out with 200 ng of the expression plasmid and 4 ng of phRL-SV40 vector (Promega) (internal control for transfection efficiency) in serum-free RPMI medium. Serum was added at a final 10% concentration after an overnight reaction. After a 24-h transfection period, we collected cells and measured luciferase activity using the Dual-Luciferase Reporter Assay System (Promega).

For interferon activation, we incubated the transfected cells with 10 ng ml^{–1} of IFN- γ (R&D systems) for another 24 h. Promoter activity is expressed as the ratio of firefly to *Renilla* luciferase values.

For co-transfection experiments, we used the *AIRE*-expression plasmid pcDNA3.1-*AIRE*¹⁷ provided by P. Peterson. Increasing doses of *AIRE* plasmid were mixed with the P1 or P3 construct and with appropriate amounts of an irrelevant plasmid to keep the total amount of transfected DNA constant. Mann–Whitney U-tests were used to assess the significance of differences between constructs.

RNA interference experiments. To investigate effects of *IRF8* on *CHRNA1* expression, we used a pool of four specific siRNA duplexes for *IRF8* designed to minimize off-target effects (On-TargetPlus Smartpool, Dharmacon; see below) as well as non-targeting duplexes, RISC-free duplexes (negative controls) and firefly-luciferase-targeting duplexes (positive control). Short interfering RNAs were co-transfected with the different constructs at a final concentration of 100 nM, as recommended by the manufacturer. Luciferase activity was assayed as above after a 48-h transfection period. Inhibition of *IRF8* expression was monitored by immunoblot analysis of cells extracts, using anti-IRF8 antibody (C19, Santa-Cruz Biotechnology) and a peroxidase-coupled secondary antibody with ECL-plus luminescent substrate (Amersham), showing a 37% average decrease. The sequences of the sense and anti-sense oligonucleotides forming the four duplexes of the anti-IRF8 pool are: duplex 1, 5'-GAGGUUACGCGUGUGCUUUGUU-3' and 5'-CAAAGCACAGCGUAAACUCUU-3'; duplex 2, 5'-CCGGAUGUUUCCAGAUUUUU-3' and 5'-AAUUCUGGAAACA-UCCGGUU-3'; duplex 3, 5'-CAUACAAGUUUACCGAAUUU-3' and 5'-AUUCGGUAAACUUUGUAUGUU-3'; duplex 4, 5'-AGCCUUCUGUGGACGAUUAUU-3' and 5'-UAAUCGUCCACAGAGGCUUU-3'.

ChIP assay. ChIP was performed using the ChIP-IT kit (Active Motif) according to the manufacturer's instructions with slight modifications. Briefly, cells were fixed with 1% formaldehyde for 30 min at room temperature. After neutralization, cells were lysed and the extracts were sonicated (20 s each, 10 times at 25% power using a Vibracell VC100 sonicator) to shear the chromatin into 200–1,000-bp fragments. The resulting sheared chromatin was pre-cleared with salmon sperm DNA/protein G agarose beads and incubated with 5 μ g of normal goat immunoglobulin G (IgG) or goat anti-IRF8 antibody (C-19, Santa-Cruz Biotechnology) overnight at 4 °C with rotation. Before incubation, a fraction of the pre-cleared chromatin was saved as an input control. Antibody/chromatin complexes were precipitated by addition of salmon sperm DNA/protein G agarose beads and eluted after extensive washing of the beads. Immunoprecipitated DNA was purified following reversal of cross-linking and digestion of RNA and proteins with RNase A and proteinase K, respectively. Levels of precipitated promoter DNA (ChIP signals) were quantified by quadruplicate real-time PCR and shown as values relative to the input DNA. The following primer pairs were used (sense and anti-sense, respectively): *CHRNA1* promoter 5'-CTGAAATCCAGGCAGTCAGA-3' and 5'-CTGCAACATGGCTTCAGTT-3'; *GAPDH* promoter (irrelevant control), 5'-TACTAGCGGTTTACGGGCG-3' and 5'-TCGAACAGGAGGAGCAGAGAGCGA-3'.

Isolation of hTECs, RNA isolation and cDNA synthesis. Samples of thymic tissue from 25 patients (7 females and 18 males; age range from 3 days to 3 yr) were obtained in the course of corrective cardiac surgery at the Department of Cardiac Surgery, Medical School, University of Heidelberg. None of the patients exhibited symptoms of an autoimmune disorder at the time undergoing treatment. This study has been approved by the Ethics Committee of the Medical School of the University of Heidelberg.

Thymic epithelial cell (TEC) purification, RNA isolation and cDNA synthesis were performed as described previously²⁸ with the following modifications. TECs were enriched on a discontinuous Percoll density gradient after sequential digestion of thymic tissue by collagenase-dispase and trypsin. Typical flow cytometry plots used for sorting TECs are shown in Supplementary Fig. 8. The cDNA was purified by centrifugation through MicroSpinG-50 columns (Amersham).

Quantitative PCR. Real time PCR reactions were performed in final volumes of 25 μ l with the optimal concentration of the forward and reverse primers (50–900 nM) using the qPCR core kit for SybrGreen I (Eurogentec) containing Hot GoldStar polymerase and uracil-DNA glycosylase. Reactions were run on a

sequence detection system (GeneAmp 5700; Applied Biosystems) in triplicate and expression values were normalized to *GAPDH* expression using the comparative C_T method. Each measurement was done twice. Primers were purchased from MWG Biotech AG, and when possible, were designed to span at least one intron. The following primer pairs were used (sense and anti-sense, respectively): *AIRE*, 5'-GAACGGGATTTCAGACCATGT-3' and 5'-TCTTCGAACCTTGCTGG-GAGT-3'; *CHRNA1*, 5'-TCGTCACCACTTCCCTTT-3' and 5'-CCGCTC-TCCATGAAGTTGCT-3'; *GAPDH*, 5'-TCGACAGTCAGCCGCATCT-3' and 5'-CCGTTGACTCCGACCTTCA-3'; *IRF8*, 5'-CGAGCCATACAAAGTTTAC-CGAAT-3' and 5'-TCACGCAGCCAGCAGTTG-3'.

Statistical analysis. The age distributions at disease onset were normalized by square root transformation (Shapiro–Wilk's W test not significant). The effect of the 11 tagSNPs on this quantitative trait was tested in the French cohort by forward stepwise multiple linear regression. The influence of rs16862847 on this same trait in the UK cohort was tested by a Student's t -test. Homogeneity of variances in patients with or without the risk allele was verified (Levene's test and Brown–Forsythe test not significant). We also ruled out a bias by potential outliers by removing the upper and lower five percentiles of the distribution, altogether making the finding robust. Calculations of the statistics and of P -values were carried out with the Statistica 6.0 software (Statsoft Inc.). Odds ratios, their 95% confidence interval (CI), the exact P -values of 2×2 tables and stratified 2×2 tables corresponding to the two cohorts of patients were calculated with the StatXact software v7 (Cytel corp).

The analysis of quantitative PCR data was based on ΔC_T values. We tested a correlation between *AIRE* and *CHRNA1* message levels with the Spearman rank R coefficient. The linear model analysis with the factors rs16862847 and *AIRE* was performed with PROC GLM of SAS version 8.2 (SAS Institute Inc.). Variances of *CHRNA1* residuals after removal of *AIRE*-correlated variations were homogeneous in rs16862847 genotypic groups (AA versus AG+GG).

26. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
27. Carlson, C. S. et al. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.* **74**, 106–120 (2004).
28. Gotter, J., Brors, B., Hergenbahn, M. & Kyewski, B. Medullary epithelial cells of the human thymus express a highly diverse selection of tissue-specific genes colocalized in chromosomal clusters. *J. Exp. Med.* **199**, 155–166 (2004).

LETTERS

A central integrator of transcription networks in plant stress and energy signalling

Elena Baena-González^{1*}, Filip Rolland^{1,2,3*}, Johan M. Thevelein^{2,3} & Jen Sheen¹

Photosynthetic plants are the principal solar energy converter sustaining life on Earth. Despite its fundamental importance, little is known about how plants sense and adapt to darkness in the daily light–dark cycle, or how they adapt to unpredictable environmental stresses that compromise photosynthesis and respiration and deplete energy supplies. Current models emphasize diverse stress perception and signalling mechanisms^{1,2}. Using a combination of cellular and systems screens, we show here that the evolutionarily conserved *Arabidopsis thaliana* protein kinases, KIN10 and KIN11 (also known as AKIN10/At3g01090 and AKIN11/At3g29160, respectively), control convergent reprogramming of transcription in response to seemingly unrelated darkness, sugar and stress conditions. Sensing and signalling deprivation of sugar and energy, KIN10 targets a remarkably broad array of genes that orchestrate transcription networks, promote catabolism and suppress anabolism. Specific bZIP transcription factors partially mediate primary KIN10 signalling. Transgenic KIN10 overexpression confers enhanced starvation tolerance and lifespan extension, and alters architecture and developmental transitions. Significantly, double *kin10 kin11* deficiency abrogates the transcriptional switch in darkness and stress signalling, and impairs starch mobilization at night and growth. These studies uncover surprisingly pivotal roles of KIN10/11 in linking stress, sugar and developmental signals to globally regulate plant metabolism, energy balance, growth and survival. In contrast to the prevailing view that sucrose activates plant SnRK1s (Snf1-related protein kinases)^{3–6}, our functional analyses of *Arabidopsis* KIN10/11 provide compelling evidence that SnRK1s are inactivated by sugars and share central roles with the orthologous yeast Snf1 and mammalian AMPK in energy signalling.

We surveyed the public transcriptome databases using the genome tool Genevestigator⁷ and discovered that dark-induced (*DIN*) genes are activated under diverse stress conditions and repressed by sugars and light (Supplementary Fig. 2a). Consistently, quantitative analyses confirmed that *DIN1* (At4g35770) and *DIN6* (At3g47340) were activated by various stresses—such as darkness, herbicide (3-(3,4-dichlorophenyl)-1,1-dimethylurea, DCMU)⁸ and flooding/submergence—that limit photosynthesis and respiration, and hence sugar and energy supplies in isolated mesophyll cells and whole plants. This activation was antagonized by both glucose and sucrose (Fig. 1a–c, and Supplementary Fig. 2b). The results suggest that plant cells sense sugar or energy depletion in a cell-autonomous manner and trigger convergent gene activation in response to diverse and seemingly unrelated stress signals independent of light.

We have previously shown that the *Arabidopsis* glucose sensor hexokinase 1 (HXK1) is essential for the glucose repression of photosynthesis genes⁹. However, *DIN* gene repression by glucose in

the dark was still active in the *HXK1*-null mutant *gin2* (*glucose-insensitive2*) (Fig. 1c), indicating the involvement of an *HXK1*-independent signalling pathway¹⁰. To identify the key regulators governing this stress/sugar signalling pathway, we developed a cell

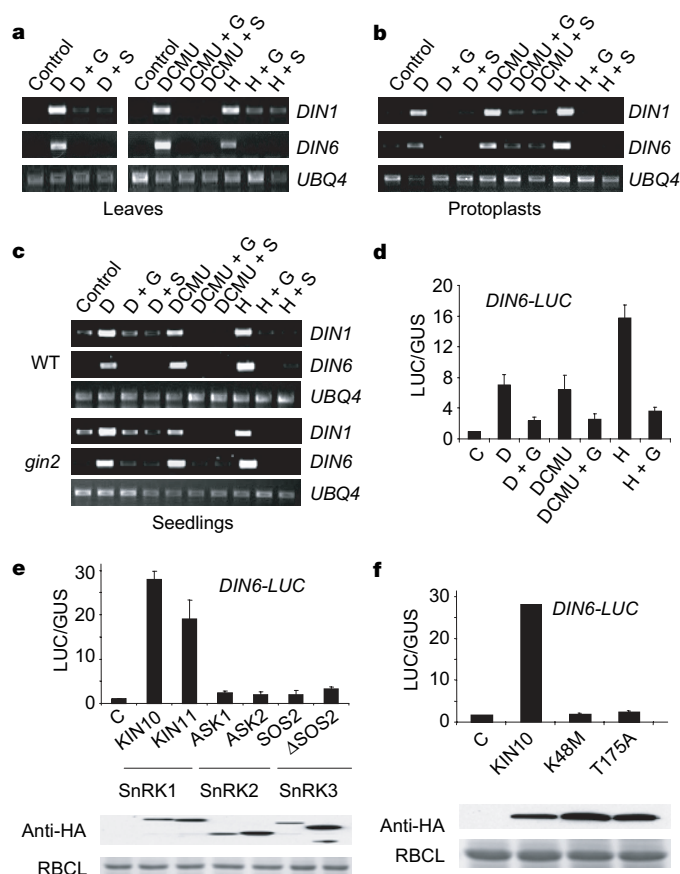


Figure 1 | *DIN* genes are activated by diverse stresses and repressed by sugars. Stress induces *DIN* gene expression in *Arabidopsis* leaves (a), and isolated mesophyll cells (b). The induction is diminished by glucose (G) and sucrose (S). c, The *gin2* mutant exhibits normal *DIN* gene induction and sugar repression. d, Diverse stresses activate the *DIN6-LUC* reporter. Glucose antagonizes *DIN6-LUC* activation. e, *DIN6-LUC* is specifically activated by SnRK1s. f, *DIN6-LUC* induction requires KIN10 kinase activity (K48) and T-loop phosphorylation (T175). D, dark; H, hypoxia; C, control DNA. Control, DCMU and hypoxia treatments were performed under light (50 $\mu\text{mol m}^{-2}\text{s}^{-1}$). Error bars, s.d. ($n = 3$). Protein expression was confirmed by western blot analysis.

¹Department of Genetics, Harvard Medical School, and Department of Molecular Biology, Massachusetts General Hospital, Boston Massachusetts 02114, USA. ²Department of Molecular Microbiology, VIB, B-3001 Leuven-Heverlee, Flanders, Belgium. ³Laboratory of Molecular Cell Biology, Institute of Botany and Microbiology, K.U.Leuven, Kasteelpark Arenberg 31, B-3001 Leuven-Heverlee, Flanders, Belgium.

*These authors contributed equally to this work.

model system using a sensitive reporter, generated by fusing the putative promoter of *Arabidopsis* *DIN6* (encoding the glutamine-dependent asparagine synthetase, *ASN1*) to the luciferase (*LUC*) gene (Methods Summary). In transfected mesophyll protoplasts, *DIN6-LUC* was activated by darkness, DCMU, hypoxia/submergence and other stresses within 3–6 h (Fig. 1d, and J.S., unpublished), in a similar way to the endogenous *DIN6* gene^{8,11–13} (Fig. 1a–c; and Supplementary Fig. 2a). This activation was repressed by sucrose or glucose at physiological concentrations as low as 5 mM (Supplementary Fig. 2b). Interestingly, endogenous *DIN6* activation by diverse signals was abolished by the protein kinase inhibitor K252a (Supplementary Fig. 2c), suggesting the requirement of protein kinases in the integration of stress and sugar signals.

We first investigated the involvement of members of the Snf1-related kinase (SnRK) gene families³ (Supplementary Fig. 3a), implicated in metabolic regulation owing to their sequence homology to the yeast Snf1 and mammalian AMPKs^{3,14,15}. Cell-based functional screens revealed the specific activation of *DIN6-LUC* by the two ubiquitously expressed members of the SnRK1 group, KIN10 and KIN11 (Fig. 1e, and Supplementary Fig. 3b, c). Representative SnRK2 and SnRK3 members lacked the same function (Fig. 1e) despite their established protein kinase activities in osmotic and salt stress responses (Methods Summary). Expression of KIN10 and KIN11 in protoplasts conferred significant kinase activity, which required the conserved ATP binding site (K48 and K49) and T175 (ref. 14; Supplementary Fig. 4, and Fig. 1f). However, unlike T172 phosphorylation in AMPK, T175 phosphorylation in KIN10 was not correlated with activation by dark, DCMU or hypoxia treatment (data not shown). Thus, *DIN6-LUC* served as a sensitive and quantitative transcription reporter to monitor endogenous KIN10/11 activity *in vivo*.

To identify critical DNA sequences involved in the stress and KIN10-mediated responses, we performed systematic mutagenesis of predicted *cis*-regulatory elements (Supplementary Methods) in the *DIN6* promoter (Supplementary Fig. 5, and Fig. 2a). Significantly, specific mutation of the G-box (CACGTG, G1) proximal to the TATA box abolished most of the *DIN6-LUC* activation by KIN10, hypoxia, darkness and DCMU (Fig. 2b–d). The results indicate that a common *cis*-element mediates convergence of diverse signals, most probably through KIN10.

To substantiate this finding further, we screened *Arabidopsis* bZIP transcription factors¹⁶ that have been characterized as G-box binding factors (GBFs), and found that GBF5/bZIP2 (At2g18160) specifically activated *DIN6-LUC* (Fig. 2e). Importantly, co-expression of KIN10 with GBF5 resulted in a dramatic and synergistic effect on *DIN6-LUC* activation. Both the GBF5 effect and its synergism with KIN10 were abolished by the specific *DIN6* promoter mutation in the G-box *g1* (Fig. 2f). Similar results were obtained with closely related bZIP S-group members expressed in mesophyll cells (bZIP11/ATB2, At4g34590; bZIP53, At3g62420; and related bZIP1, At5g49450; Supplementary Table 1 and Fig. 2g, h). Functional redundancy of the transcription factors was evident also from the lack of overt phenotypes in the available loss-of-function single *bzip* mutants (data not shown). Because the S- and C-group bZIPs have been shown to form functional heterodimers¹⁶, we also examined the activity of C-group bZIPs. Interestingly, one member of the C-group, bZIP63, showed little activity alone but exhibited moderate synergism with KIN10 through the G-box (Fig. 2h). These studies identified novel functions of specific GBFs/bZIPs and provide further evidence for convergent responses in stress signalling through KIN10 and KIN11.

To explore the extent of KIN10 transcriptional regulation and identify its early target genes, we scaled-up transient protoplast expression experiments, which could rule out secondary or long-term effects of metabolism and growth, and circumvent limitations caused by redundancy and embryonic lethality observed in mammals and plants^{15,17,18}. Gene expression profiles with or without KIN10

expression were compared using *Arabidopsis* whole-genome ATH1 GeneChips (GEO accession number GSE8257; Supplementary Methods). To filter candidate KIN10 target genes (at least twofold change with *P*value < 0.0004), we combined independent experimental validation using quantitative real-time reverse transcriptase PCR (qRT-PCR) with vigorous statistical analyses of GeneChip data from biological replicates, using parametric and non-parametric methods (see Supplementary Methods and Supplementary Fig. 6 for details). Guided by the experimentally validated MYB75 transcription factor gene and 24 other marker genes, we were able to define 1,021 putative KIN10 target genes that exhibited reproducible activation or repression (Supplementary Tables 1–3).

To increase robustness and assess the physiological significance of the co-regulation of KIN10 target genes, we performed additional computational screens to identify seven highly correlated gene expression patterns from published data sets generated with the *Arabidopsis* ATH1 GeneChips^{11–13,19–22} (details in Supplementary

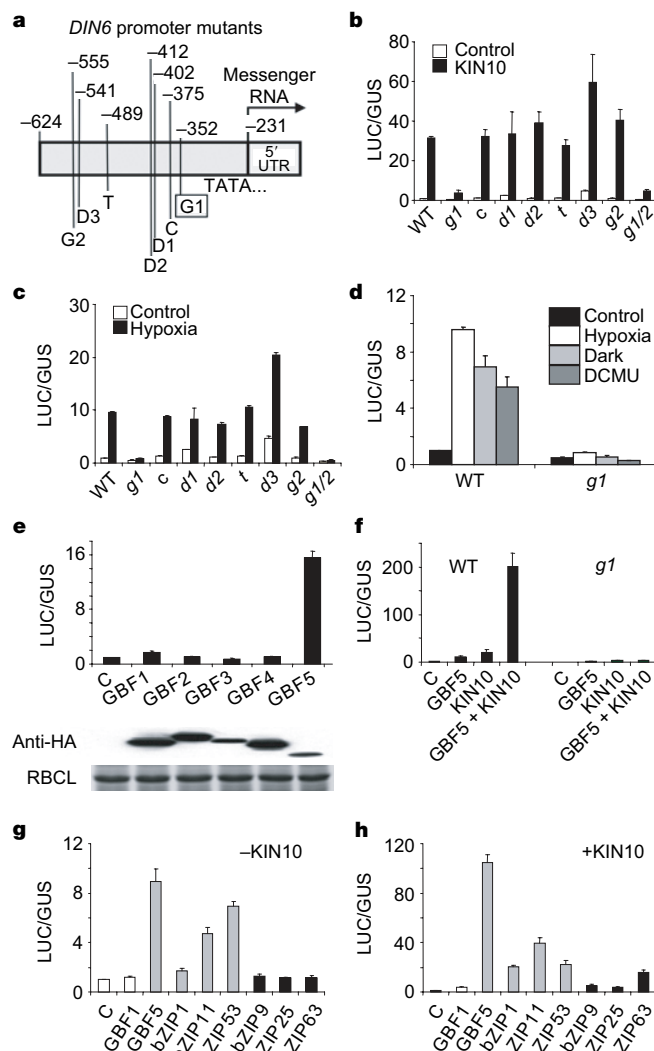


Figure 2 | Convergent transcriptional control through a G-box and specific GBFs. **a–d**, Mutational analysis of the *DIN6* promoter. G, G-box; T, TATCCA element; D, DOF-binding *cis*-element; C, C-box/ACGT/OSAMY3. Numbers are relative to ATG. The *g1* mutation prevents activation by KIN10 (**b**), hypoxia (**c**), dark and DCMU (**d**). **e**, *DIN6-LUC* is specifically induced by GBF5. **f**, KIN10 and GBF5 synergistically activate *DIN6-LUC* through the G1 box. **g**, Functional redundancy of S-group bZIPs (grey) in *DIN6-LUC* activation. **h**, Synergistic activation of *DIN6-LUC* by KIN10 and S-group (grey) and C-group (black; bZIP63) bZIPs. WT, wild-type *DIN6* promoter; C, control DNA. Control, DCMU, and hypoxia treatments were performed under light. Error bars, s.d. (*n* = 3).

Methods). Specifically, gene expression profiles generated under various sugar and energy starvation conditions^{11,20–22} revealed a striking positive correlation with KIN10 target genes (Pearson correlation coefficient 0.85 to 0.87, Supplementary Table 5). Importantly, KIN10 target genes also exhibited strong negative correlations (Pearson correlation coefficient -0.87 to -0.92 , Supplementary Table 5) with the gene expression profiles obtained from glucose- or sucrose-treated seedlings and differential CO₂-fixing adult leaves in intact plants^{12,13,19}.

A final filtering step was applied to remove genes with inconsistent expression patterns in any of the seven data sets. The stringent and multi-step filtering processes selected a reliable list of 278 genes co-activated by KIN10 and sugar starvation conditions, and co-repressed in sugar-treated seedlings or CO₂-fixing adult leaves (Fig. 3a, and Supplementary Table 4). A second list of 322 genes was also identified on the basis of their co-repression by KIN10 and sugar starvation conditions, but co-activation in sugar-treated seedlings or differential CO₂-fixing adult leaves (Fig. 3b, and Supplementary Table 4). As predicted by the *DIN* gene response (Fig. 1), hypoxia-treated protoplasts also displayed gene expression profiles similar to KIN10 overexpression on the basis of ATH1 GeneChip (GEO accession number GSE8248) and triplicated analyses of selected marker genes by qRT-PCR (Supplementary Fig. 7a, b). The correlation across the data sets (Supplementary Table 5) was remarkable considering the diversity of materials and experimental conditions used for the comparison and filtering (see Supplementary Methods). This suggests that KIN10 regulation is a cell-autonomous and ubiquitous event and highlights its physiological relevance.

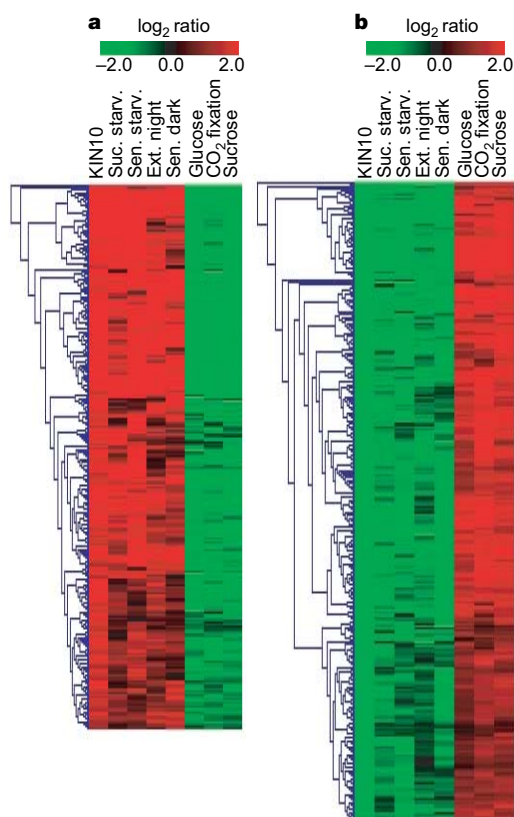


Figure 3 | Global gene expression regulation by KIN10. **a, b,** Transient KIN10 expression in protoplasts results in the induction (**a**) and repression (**b**) of genes involved in a wide variety of cellular processes and metabolism (Supplementary Fig. 7c–e, and Supplementary Tables 3, 4). The KIN10-mediated gene expression profile is positively correlated with that regulated by sugar and carbon deprivation, and negatively correlated with that controlled by sugars (see text for details). Ext., extended; sen., senescence; starv., starvation; suc., sucrose.

Previous studies have identified specific SnRK1 target genes, one each in potato and wheat^{3,5}. The present work provides for the first time a complete overview of the surprisingly broad genome-wide transcript changes induced by these conserved protein kinases in a multicellular organism. The most prominent KIN10-activated (and sugar-repressed) genes represented a variety of major catabolic pathways, including cell wall, starch, sucrose, amino acid, lipid, and protein degradation that provide alternative sources of energy and metabolites (Supplementary Fig. 7c, and Supplementary Tables 3, 4). KIN10 also induced several APG/ATG (AUTOPHAGY) orthologues and plant-specific trehalose metabolism genes (*TPS8-11*), which potentially alter or sense the level of trehalose-6-P, a regulator of plant carbohydrate metabolism, growth and development^{23,24}. Conversely, a large set of genes involved in energy-consuming ribosome biogenesis (87 genes) and anabolism (Supplementary Fig. 7d, e) were co-ordinately repressed by KIN10 (Supplementary Tables 3, 4). Although many plant metabolic pathways have been well characterized at the biochemical level, their molecular regulatory mechanisms are still poorly understood. Thus, besides the well-known roles of SnRK1 in phosphorylating and modulating enzymes important for carbon and nitrogen metabolism^{3,14,25,26}, the present work identifies KIN10 and KIN11 as global regulators of gene expression involved in primary and secondary metabolism and protein synthesis^{11–13,19–22,27}. The results also reveal central and previously unrecognized regulatory roles of the plant SnRK1 (and possibly also mammalian AMPK) in controlling a large number of genes encoding transcription factors, chromatin remodelling factors, and a plethora of signal transduction components (Supplementary Fig. 7c, e, and Supplementary Tables 3, 4).

Significantly, many KIN10 marker genes induced by multiple stresses are primary glucose repressible genes insensitive to cycloheximide¹² and direct targets of GBF5 (Supplementary Fig. 7f). These genes are not induced by the inactive KIN10 K48M or T175A mutant (data not shown), pointing to a requirement of KIN10 protein kinase activity for this global gene expression response. Our results suggest that KIN10/11 initiate multiple transcription cascades in response to sugar or energy depletion in darkness and stress conditions. Coordinated transcription activation of direct target genes is at least in part exerted through GBF5 and related bZIP transcription factors. Future analyses will aim at elucidating the precise molecular link between the protein kinases and transcription factors.

To study the long-term effects of KIN10 activity at the whole-plant level, we generated *Arabidopsis* transgenic plants overexpressing KIN10 (*O1* and *O2*) or with reduced KIN10 expression using RNA interference (RNAi; *10-1* and *10-2*) (Supplementary Fig. 8a). The KIN10 overexpression seedlings displayed some advantage in primary root growth and development under low light with limited energy supply (Supplementary Fig. 8b, e). With exogenous sucrose, KIN10 overexpression plants showed reduced growth in shoots and roots, possibly owing to KIN10 repression of biosynthetic activities (Supplementary Fig. 8c–e). In contrast, the KIN10 silenced lines were able to use exogenously supplied sucrose (Supplementary Fig. 8c–e) and glucose (data not shown) more efficiently and exhibited enhanced shoot and root growth. Growth on 3% sucrose also increased anthocyanin accumulation in wild-type, *10-1* and *10-2* plants, but not in the KIN10 overexpression lines (Supplementary Fig. 8f). Consistently, KIN10 overexpression repressed the expression of *MYB75/PAP1* (At1g56650) (Supplementary Table 3), a key transcription factor for anthocyanin biosynthesis (Supplementary Methods).

A role of KIN10 in the starvation response was most evident after transferring seedlings to glucose-free liquid medium under a low light intensity that limits photosynthesis. Whereas wild-type seedlings underwent rapid senescence on nutrient deprivation, KIN10 overexpression promoted plant survival (Fig. 4a). When grown in soil, KIN10 overexpression altered inflorescence architecture (Fig. 4b, c) and delayed flowering and onset of senescence under long-day

conditions (20 h light/4 h dark) (Fig. 4c). This is reminiscent of life-span extension by calorie restriction in animals and provides evidence for a new role of KIN10 in determining plant shape and developmental transition timing and linking this to metabolism and environmental cues. Soil-grown *10-1* and *10-2* lines did not show an obvious phenotype, probably owing to functional redundancy of KIN11.

To circumvent the experimental limitation of obtaining double loss-of-function mutants^{3,17,18}, we generated *kin11* single- and *kin10 kin11* double-mutant plants, using virus-induced gene silencing (VIGS)²⁸ of *KIN11* in wild-type and *10-1* or *10-2* transgenic seedlings. Reduction of *KIN* transcripts was confirmed by RT-PCR and immunoblotting (using KIN10 and P-AMPK antibodies; Supplementary Fig. 9a) before extensive phenotypic and molecular studies. Three weeks after infiltration, double mutants *d-1* and *d-2* showed a surprisingly dramatic growth defect with small leaves and short petioles under 13 h light/11 h dark (Fig. 4d, and Supplementary Fig. 9b) or constant illumination conditions (data not shown). A strong accumulation of anthocyanins, curling of new leaves (leaves 11–12) and symptoms of early senescence were visible (Fig. 4d). The impact on growth and senescence in *d-1* and *d-2* lines was most striking five weeks after infiltration, when the transition to the reproductive phase occurred in wild-type plants (Fig. 4e). The most severely silenced plants senesced before flowering (Fig. 4e), whereas those with milder silencing bolted but failed to produce viable primary inflorescences and axillary floral meristems (Fig. 4f). Unlike the *snf1a snf1b* double mutant in moss²⁹, the growth phenotypes could neither be rescued by continuous light irradiance nor by supplementation with 1% sucrose (data not shown). This argues against the growth defects being merely due to impaired catabolism, and supports a more fundamental and unexpected role of

KIN10/KIN11 in normal vegetative and reproductive growth and development in flowering plants.

To establish a definitive molecular and quantitative link between KIN10/11 action and the ability to mount transcription activation in response to stresses and energy deprivation, we examined the response to hypoxia, dark or DCMU in mesophyll cells and intact leaves of wild-type, single- and double-mutant plants. All KIN10 marker genes tested could no longer be activated by any of these stresses in the *d-1* and *d-2* double mutants (Fig. 4g, h; Supplementary Fig. 9c, d, and Supplementary Table 6). Consistent with a role of KIN10/11 in promoting catabolic processes, leaf starch—as a major carbon source in the absence of photosynthesis—remained high at the end of night in the *d-1* and *d-2* double mutants (Fig. 4i). This suggests that KIN10/11 may have a previously unrecognized regulatory role in starch mobilization at night³⁰. Alternatively, starch accumulation may reflect a reduced energy demand for growth. In the future, it will be interesting and informative to extend the analysis of gene expression reprogramming to changes in enzyme activities and metabolite homeostasis in response to environmental cues associated with KIN10/KIN11 regulation.

Our studies provide molecular, biochemical, genetic and genomic evidence for the essential functions of KIN10/KIN11 in plant protection and survival under stress, darkness and sugar deprivation conditions. Surprising roles of KIN10/KIN11 in vegetative and reproductive growth and developmental transition under normal conditions are uncovered (Supplementary Fig. 1). Their roles as energy sensors and integrators, orchestrating global transcription, may also be found in mammals and thus our findings in plants may contribute to the understanding of AMPK functions in relation to diabetes, cancer, obesity and longevity. Further dissection of the components and molecular mechanisms of the KIN10/KIN11

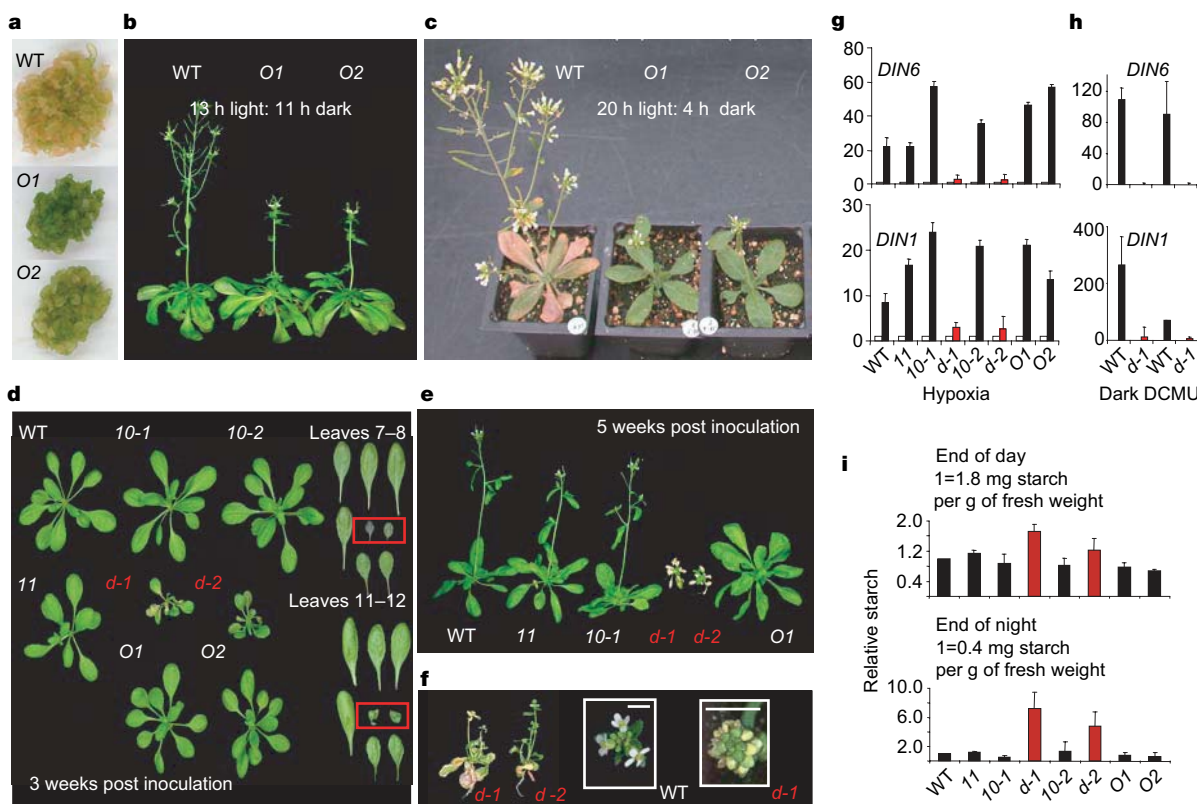


Figure 4 | Role of KIN10 in plant development, stress responses and starch mobilization at night. **a–c**, *KIN10* overexpression (*O1*, *O2*) prevents nutrient-deprivation-induced senescence (**a**), alters inflorescence development (**b**, **c**), and delays long-day-triggered senescence (**c**). **d–f**, *kin10 kin11* VIGS double mutant (*d-1*, *d-2*) plants show growth arrest (**d**) and premature senescence that precludes flowering (**e**), or bolting without

axillary floral meristems or viable inflorescences (inset) (**f**). The KIN10 target gene response to hypoxia (**g**, mesophyll cells) and dark or DCMU (**h**, leaves) is abolished in the double mutants (red). **i**, The double mutants are impaired in starch mobilization. *10-1/10-2*, *kin10* RNAi lines; *11*, *kin11* VIGS plants. Error bars, s.d. (**g**, **i**, $n = 3$; **h**, $n = 2$).

signalling cascades will provide valuable information on the metabolic control of plant growth and development, which may enable a more targeted genetic modification of plant development, architecture, carbon allocation, and stress resistance—all major determinants of crop yield and renewable energy production.

METHODS SUMMARY

Primers. All primers used are listed in Supplementary Table 7.

Effector and reporter constructs. *KIN10*, *KIN11*, *ASK1*, *ASK2* and the full-length and truncated *SOS2* (*ASOS2*, encoding amino acids 1–308) were fused to the haemagglutinin (HA) tag and cloned between a 35S-derived promoter and NOS terminator. The 624-bp promoter/5'-UTR sequence of *DIN6* was fused to the *LUC* reporter gene to generate *DIN6-LUC*.

Protoplast transient expression assay. Protoplast assays were performed using *UBQ10-GUS* as a transfection control. Protoplasts ($1-4 \times 10^4$) were incubated for 6 h in 1 ml of mannitol buffer in 6-well plates, or submerged in 1 ml of buffer in a 1.5 ml microfuge tube for hypoxia treatment. Protein expression was analysed using KIN10-specific rabbit antibody generated against a glutathione S-transferase (GST)-fused KIN10 polypeptide (amino acids 335–388), human anti-P-AMPK (New England Biolabs), and a monoclonal anti-HA antibody (Roche). For dark treatment, plates were covered with aluminium foil. DCMU was added at 20 μ M. For global gene expression analyses, transfection experiments were scaled up 50-fold. For details on GeneChip experiments, data processing and analysis, see Supplementary Information.

Protein kinase assays. These were performed using immunoprecipitated HA-tagged kinases and SPS peptide (RDHMPRIKSEMQUIWSED) as a substrate.

Transgenics. To generate *KIN10* overexpression lines, the coding region was cloned in a pCB302-derived minibinary expression vector. For *KIN10* RNAi, the last 238 bp of the coding region (primers: 5'-cgggatccagcagcgagatggtatg-3', 5'-aaggctcatgcatggtcagagactcggagctgag-3') were cloned in both orientations, separated by an intron.

Virus-induced gene silencing (VIGS). *KIN11* VIGS was performed as described²⁸ using a gene-specific 503-bp *KIN11* fragment (–103 to +400, primers: 5'-ggaattcgttctgtatattcttcgctc-3', 5'-cgggatccagctactctaccagatattat-3') and repeated six times with consistent results.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 April; accepted 2 July 2007.

Published online 1 August 2007.

1. Yamaguchi-Shinozaki, K. & Shinozaki, K. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.* **57**, 781–803 (2006).
2. Hasegawa, P. M., Bressan, R. A., Zhu, J.-K. & Bohnert, H. J. Plant cellular and molecular responses to high salinity. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **57**, 463–499 (2000).
3. Halford, N. G. *et al.* Metabolic signalling and carbon partitioning: role of Snf1-related (SnRK1) protein kinase. *J. Exp. Bot.* **54**, 467–475 (2003).
4. Bhalerao, R. P. *et al.* Regulatory interaction of PRL1 WD protein with *Arabidopsis* SNF1-like protein kinases. *Proc. Natl Acad. Sci. USA* **96**, 5322–5327 (1999).
5. Purcell, P. C., Smith, A. M. & Halford, N. G. Antisense expression of a sucrose non-fermenting-1-related protein kinase sequence in potato results in decreased expression of sucrose synthase in tubers and loss of sucrose-inducibility of sucrose synthase transcripts in leaves. *Plant J.* **14**, 195–202 (1998).
6. Tiessen, A. *et al.* Evidence that SNF1-related kinase and hexokinase are involved in separate sugar-signalling pathways modulating post-translational redox activation of ADP-glucose pyrophosphorylase in potato tubers. *Plant J.* **35**, 490–500 (2003).
7. Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. & Gruissem, W. GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol.* **136**, 2621–2632 (2004).
8. Fujiki, Y. *et al.* Dark-inducible genes from *Arabidopsis thaliana* are associated with leaf senescence and repressed by sugars. *Physiol. Plant.* **111**, 345–352 (2001).
9. Moore, B. *et al.* Role of the *Arabidopsis* glucose sensor HXK1 in nutrient, light, and hormonal signaling. *Science* **300**, 332–336 (2003).
10. Rolland, F., Baena-González, E. & Sheen, J. Sugar sensing and signaling in plants: conserved and novel Mechanisms. *Annu. Rev. Plant Biol.* **57**, 675–709 (2006).
11. Lin, J. F. & Wu, S. H. Molecular events in senescing *Arabidopsis* leaves. *Plant J.* **39**, 612–628 (2004).
12. Price, J., Laxmi, A., St Martin, S. K. & Jang, J. C. Global transcription profiling reveals multiple sugar signal transduction mechanisms in *Arabidopsis*. *Plant Cell* **16**, 2128–2150 (2004).
13. Palenchar, P. M., Kouranov, A., Lejay, L. V. & Coruzzi, G. M. Genome-wide patterns of carbon and nitrogen regulation of gene expression validate the

combined carbon and nitrogen (CN)-signaling hypothesis in plants. *Genome Biol.* **5**, R91 (2004).

14. Hardie, D. G., Carling, D. & Carlson, M. The AMP-activated/SNF1 protein kinase subfamily: metabolic sensors of the eukaryotic cell? *Annu. Rev. Biochem.* **67**, 821–855 (1998).
15. Kahn, B. B., Alquier, T., Carling, D. & Hardie, D. G. AMP-activated protein kinase: ancient energy gauge provides clues to modern understanding of metabolism. *Cell Metab.* **1**, 15–25 (2005).
16. Jakoby, M. *et al.* bZIP transcription factors in *Arabidopsis*. *Trends Plant Sci.* **7**, 106–111 (2002).
17. Zhang, Y. *et al.* Expression of antisense SnRK1 protein kinase sequence causes abnormal pollen development and male sterility in transgenic barley. *Plant J.* **28**, 431–441 (2001).
18. Radchuk, R., Radchuk, V., Weschke, W., Borisjuk, L. & Weber, H. Repressing the expression of the *SUCROSE NONFERMENTING-1-RELATED PROTEIN KINASE* gene in pea embryo causes pleiotropic defects of maturation similar to an abscisic acid-insensitive phenotype. *Plant Physiol.* **140**, 263–278 (2006).
19. Bläsing, O. E. *et al.* Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. *Plant Cell* **17**, 3257–3281 (2005).
20. Buchanan-Wollaston, V. *et al.* Comparative transcriptome analysis reveals significant differences in gene expression and signalling pathways between developmental and dark/starvation-induced senescence in *Arabidopsis*. *Plant J.* **42**, 567–585 (2005).
21. Contento, A. L., Kim, S. J. & Bassham, D. C. Transcriptome profiling of the response of *Arabidopsis* suspension culture cells to Suc starvation. *Plant Physiol.* **135**, 2330–2347 (2004).
22. Thimm, O. *et al.* MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**, 914–939 (2004).
23. Schluepmann, H., Pellny, T., van Dijken, A., Smeekens, S. & Paul, M. Trehalose 6-phosphate is indispensable for carbohydrate utilization and growth in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **100**, 6849–6854 (2003).
24. Gomez, L. D., Baud, S., Gilday, A., Li, Y. & Graham, I. A. Delayed embryo development in the *ARABIDOPSIS* TREHALOSE-6-PHOSPHATE SYNTHASE 1 mutant is associated with altered cell wall structure, decreased cell division and starch accumulation. *Plant J.* **46**, 69–84 (2006).
25. Sugden, C., Donaghy, P. G., Halford, N. G. & Hardie, D. G. Two SNF1-related protein kinases from spinach leaf phosphorylate and inactivate 3-hydroxy-3-methylglutaryl-coenzyme A reductase, nitrate reductase, and sucrose phosphate synthase *in vitro*. *Plant Physiol.* **120**, 257–274 (1999).
26. Kaiser, W. M. & Huber, S. C. Post-translational regulation of nitrate reductase: Mechanism, physiological relevance and environmental triggers. *J. Exp. Bot.* **52**, 1981–1989 (2001).
27. Li, Y. *et al.* Establishing glucose- and ABA-regulated transcription networks in *Arabidopsis* by microarray analysis and promoter classification using a Relevance Vector Machine. *Genome Res.* **16**, 414–427 (2006).
28. Burch-Smith, T. M., Schiff, M., Liu, Y. & Dinesh-Kumar, S. P. Efficient virus induced gene silencing in *Arabidopsis thaliana*. *Plant Physiol.* **142**, 21–27 (2006).
29. Thelander, M., Olsson, T. & Ronne, H. Snf1-related protein kinase 1 is needed for growth in a normal day–night light cycle. *EMBO J.* **23**, 1900–1910 (2004).
30. Smith, A. M., Zeeman, S. C. & Smith, S. M. Starch degradation. *Annu. Rev. Plant Biol.* **56**, 73–98 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A model for KIN10/11 function is available in Supplementary Information (Supplementary Fig. 1).

Acknowledgements We thank S. P. Dinesh-Kumar for generously sharing the TRV-based vectors and the VIGS protocol for *Arabidopsis* plants before publication, O. Thimm and M. Stitt for the MAPMAN program and the original functional classification files, J. C. Jang, S. Wu and D. C. Bassham for sharing the original GeneChip data, B. Wittner for consultation on the RankProd analysis, Q. Hall for the pQH29 RNAi vector, L. Zhou for the *DIN6-LUC* construct, L. Shan and P. He for the VIGS GFP control vector and the infiltration procedure, J. Bush for plant management, K. Chu for data analysis, and the Sheen laboratory members for stimulating discussions. The work was supported by grants from the National Science Foundation and National Institute of Health to J. S. F.R. was supported by a return grant from the Belgian Office for Scientific, Technical and Cultural Affairs and fellowships from the Belgian American Educational Foundation and the Research Foundation–Flanders (FWO–Vlaanderen).

Author Information All microarray data are available at the Gene Expression Array Omnibus website (<http://www.ncbi.nlm.nih.gov/geo/>) under accession numbers GSE8248 and GSE8257. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to F.R. (filip.rolland@bio.kuleuven.be) and requests for materials should be addressed to F.R. and E.B.-G. (baena@molbio.mgh.harvard.edu).

METHODS

Phylogenetic tree. For generation of the phylogenetic tree (Supplementary Fig. 3a), protein sequences were aligned with ClustalW, using the SDSC Workbench (workbench.sdsc.edu).

Effector and reporter constructs. Effector constructs—*KIN10* (SnRK1.1; At3g01090), *KIN11* (SnRK1.2; At3g29160), and *SOS2* (SnRK3.1; At5g35410)—were amplified from leaf complementary DNA, fused to the HA tag, and cloned between a 35S-derived promoter and NOS terminator^{31–33}. The *ASK1* (SnRK2.4; At1g10940) and *ASK2* (SnRK2.1; At5g08590), and the constitutively active version of *SOS2* (Δ SOS2, amino acids 1–308) were as described^{31,34,35}. Reporter construct: The 624-bp promoter/5'-UTR sequence of *DIN6* (At3g47340) was fused to the *LUC* reporter gene to generate *DIN6-LUC* as described^{32,33,36}. The kinase (K48M, K49M, T175A) and *DIN6* promoter mutants were generated by site-directed mutagenesis, as described³³.

Protoplast transient expression assay and treatments. Protoplast transient expression assays were carried out as described^{32,33,37}, using 35S-*GUS* or *UBQ10-GUS* reporter gene fusions as transfection controls^{32,33}. Data were generated from at least three independent experiments with consistent results. Protoplasts ($1\text{--}4 \times 10^4$) were incubated for 6 h in 1 ml of mannitol buffer in 6-well tissue culture plates (1 mm depth, excluding hypoxia effects)^{32,33,37} or submerged in 1 ml of mannitol buffer in a 1.5 ml microfuge tube (25-mm depth) for hypoxia treatment. Establishment of the hypoxic condition was confirmed by the induction of well-known hypoxia marker genes encoding ADH and PDC (refs. 38, 39) in qRT-PCR and microarray analyses (data not shown). These genes are induced under hypoxic and anoxic conditions, but not under other types of cellular energy stress. Unlike *DIN1* and *DIN6*, their induction cannot be repressed by sugar³⁹. For dark treatment, plates were covered with aluminium foil for 6 h. If not otherwise indicated, DCMU, glucose and sucrose, were added to final concentrations of 20 μ M, 25 mM and 50 mM, respectively.

RT-PCR analysis. The fifth and sixth leaves of 4-week-old plants, protoplasts and one-week-old seedlings were treated with darkness (leaves and protoplasts, 6 h; leaves of VIGS plants, 10 h; seedlings, 20 h), DCMU (leaves and seedlings, 100 μ M, 20 h; leaves of VIGS plants, 100 μ M, 10 h; protoplasts, 50 μ M, 6 h) or submerged in water (hypoxia treatment, leaves and seedlings, 20 h; protoplasts, 6 h) in the absence or presence of 3% sucrose, 25 mM glucose, or a protein kinase inhibitor K252a (2 μ M). RNA samples were isolated and analysed by RT-PCR for *DIN1* and *DIN6* expression using gene-specific primers (Supplementary Table 7). *UBQ4* was used as a control gene. Total RNA (1 μ g) was converted to single-stranded complementary DNA by reverse transcriptase (Invitrogen). Quantitative real time RT-PCR (qRT-PCR) was performed using PCR Master Mix SYBR Green (BioRad) in a BioRad iCycler (3 min 95 °C, 40 cycles at 10 s 95 °C, 45 s at 59 °C, and 1 min 95 °C, 1 min 55 °C). In the protoplast assays, marker gene expression was normalized to *CIPK23* (At1g30270) expression, chosen on the basis of its steady levels in the *KIN10* and hypoxia GeneChip experiments (Supplementary Table 1). In the leaf assays, marker gene expression was normalized to *TUB4* (At5g44340).

Protein kinase assays. Kinase assays were performed as described⁴⁰ using HA-tagged kinases immunoprecipitated from protoplasts and the SPS peptide (RDHMPRIKSEMQUIWSED) as a specific substrate⁴¹. Kinase reaction mixes were spotted onto Whatman P81 phosphocellulose paper squares and the reactions were terminated by immersion in 75 mM phosphoric acid. Squares were washed twice, airdried and incorporated radioactivity was measured by liquid scintillation counting.

Antibodies and protein expression analyses. To generate the *KIN10*-specific antibody, a GST-fused *KIN10* polypeptide (amino acids 335–388) was produced in *E. coli* using a pGEX-4T-1 vector and GST was removed before rabbit injection. The antibody was affinity purified using a SulfoLink matrix (Pierce). Phospho-*KIN10* and phospho-*KIN11* were detected using a commercial antibody against the human phospho-AMPK α subunit (P-AMPK, New England Biolabs). HA-tagged proteins were detected by an anti-HA monoclonal antibody (Roche) using standard techniques³³.

Transgenics. To generate transgenic *KIN10* overexpression lines, the coding region was cloned into an expression vector derived from the pCB302 minibinary vector³³. For generating the *kin10* RNAi lines, the last 238 bp of the coding region was cloned into pQH29 in both orientations separated by an intron (Q. Hall, unpublished information) and sub-cloned into the pCB302-derived expression vector³³. Multiple transgenic lines were generated by standard protocols³³, screened by RT-PCR and analysed for consistent phenotypes.

Loss-of-function transgenics. No T-DNA insertion, enhancer or gene trap knockout lines with significantly reduced *KIN10* or *KIN11* expression could

be obtained (data not shown). We also attempted to generate *kin11* single and *kin10 kin11* double mutants using the same RNAi strategy. However, despite multiple efforts and extensive transgenic screens, we were not able to recover transgenic lines, possibly because of embryonic lethality^{3,15}. Studies in barley pollen¹⁷ and pea seed¹⁸ have shown that antisense expression of *SnRK1* causes male sterility and embryo maturation defects, consistent with our inability to obtain transgenic *Arabidopsis* lines with deficiency in both *KIN10* and *KIN11*.

Virus-induced gene silencing (VIGS). VIGS was performed as described using *Arabidopsis* wild-type, *KIN10*-overexpressing, and *kin10* RNAi seedlings^{28,42–44}. The control VIGS construct was generated using a 596-bp *GFP* fragment and the *KIN11* VIGS construct, using a gene-specific 503-bp fragment (103 bp of the 5' UTR and the first 400 bp of the *KIN11* coding region). The mutant plants were identified by RT-PCR and the VIGS experiments were repeated six times with consistent results.

Plant growth and phenotype assays. Plants were grown in soil under standard 13 h light/11 h dark or 20 h light/4 h dark (long-day) conditions. For the various phenotype assays, seedlings were grown under constant light (80 μ E) in liquid medium: (1) 0.5 \times MS, 0.5% sucrose, 60 r.p.m. for 7 days for dark, DCMU, and hypoxia treatments; (2) MS, 2% glucose, 60 r.p.m. for 7 days and transferred to glucose-free medium for 2 weeks, for nutrient starvation assays; (3) MS, 3% sucrose, 60 r.p.m. for 7 days for spectrophotometric anthocyanin quantification (OD_{530}). To test the effect of sugar availability on growth, seedlings were grown for 11 d vertically on agar plates (0.1 \times MS \pm sucrose) under 16 h light/8 h dark conditions.

Starch quantification. Starch was extracted from 0.2–0.3 g (fresh weight) of leaf material and quantified using a starch assay kit (Sigma)⁴⁵.

URLs. Genevestigator is available at <http://www.genevestigator.ethz.ch/>; MAPMAN, at <http://gabi.rzpd.de/projects/MapMan/>; NASC International Affymetrix Service at <http://affymetrix.arabidopsis.info/>; and TIGR MeV at <http://www.tm4.org/mev.html>. Gene annotation information is available on the Sheen lab website at http://genetics.mgh.harvard.edu/sheenweb/search_affy.html.

31. Kovtun, Y., Chiu, W.-L., Zeng, W. & Sheen, J. Suppression of auxin signal transduction by a MAPK cascade in higher plants. *Nature* **395**, 716–720 (1998).
32. Kovtun, Y., Chiu, W.-L., Tena, G. & Sheen, J. Functional analysis of oxidative stress-activated MAPK cascade in plants. *Proc. Natl Acad. Sci. USA* **97**, 2940–2945 (2000).
33. Hwang, I. & Sheen, J. Two-component circuitry in *Arabidopsis* cytokinin signal transduction. *Nature* **413**, 383–389 (2001).
34. Guo, Y., Halfter, U., Ishitani, M. & Zhu, J. K. Molecular characterization of functional domains in the protein kinase SOS2 that is required for plant salt tolerance. *Plant Cell* **13**, 1383–1400 (2001).
35. Boudsocq, M., Barbier-Brygoo, H. & Lauriere, C. Identification of nine sucrose nonfermenting 1-related protein kinases 2 activated by hyperosmotic and saline stresses in *Arabidopsis thaliana*. *J. Biol. Chem.* **279**, 41758–41766 (2004).
36. Lam, H. M., Peng, S. S. & Coruzzi, G. M. Metabolic regulation of the gene encoding glutamine-dependent asparagine synthetase in *Arabidopsis thaliana*. *Plant Physiol.* **106**, 1347–1357 (1994).
37. Sheen, J. Signal transduction in maize and *Arabidopsis* mesophyll protoplasts. *Plant Physiol.* **127**, 1466–1475 (2001).
38. Gonzali, S. et al. The use of microarrays to study the anaerobic response in *Arabidopsis*. *Ann. Bot. (Lond.)* **96**, 661–668 (2005).
39. Loreti, E., Poggi, A., Novi, G., Alpi, A. & Perata, P. A genome-wide analysis of the effects of sucrose on gene expression in *Arabidopsis* seedlings under anoxia. *Plant Physiol.* **137**, 1130–1138 (2005).
40. Cheng, S.-H., Sheen, J., Gerrish, C. & Bolwell, G. P. Molecular identification of phenylalanine ammonia-lyase as a substrate of a specific constitutively active *Arabidopsis* CDPK expressed in maize protoplasts. *FEBS Lett.* **503**, 185–188 (2001).
41. Huang, J. Z. & Huber, S. C. Phosphorylation of synthetic peptides by a CDPK and plant SNF1-related protein kinase. Influence of proline and basic amino acid residues at selected positions. *Plant Cell Physiol.* **42**, 1079–1087 (2001).
42. Liu, Y., Schiff, M. & Dinesh-Kumar, S. P. Virus-induced gene silencing in tomato. *Plant J.* **31**, 777–786 (2002).
43. Ryu, C. M. A. n. a. n. d. A., Kang, L. & Mysore, K. S. Agrodrench: a novel and effective agroinoculation method for virus-induced gene silencing in roots and diverse Solanaceous species. *Plant J.* **40**, 322–331 (2004).
44. Burch-Smith, T. M., Anderson, J. C., Martin, G. B. & Dinesh-Kumar, S. P. Applications and advantages of virus-induced gene silencing for gene function studies in plants. *Plant J.* **39**, 734–746 (2004).
45. Kötting, O. et al. Identification of a novel enzyme required for starch metabolism in *Arabidopsis* leaves. The phosphoglucan, water dikinase. *Plant Physiol.* **137**, 242–52 (2005).

p15^{Ink4b} is a critical tumour suppressor in the absence of p16^{Ink4a}

Paul Krimpenfort^{1*}, Annemieke IJpenberg^{1†*}, Ji-Ying Song², Martin van der Valk², Martijn Nawijn¹, John Zevenhoven¹ & Anton Berns¹

The *CDKN2b–CDKN2a* locus on chromosome 9p21 in human (chromosome 4 in mouse) is frequently lost in cancer. The locus encodes three cell cycle inhibitory proteins: p15^{Ink4b} encoded by *CDKN2b*, p16^{Ink4a} encoded by *CDKN2a* and p14^{ARF} (p19^{Arf} in mice) encoded by an alternative reading frame of *CDKN2a* (ref. 1). Whereas the tumour suppressor functions for p16^{Ink4a} and p14^{ARF} have been firmly established, the role of p15^{Ink4b} remains ambiguous. However, many 9p21 deletions also remove *CDKN2b*, so we hypothesized a synergistic effect of the combined deficiency for p15^{Ink4b}, p14^{ARF} and p16^{Ink4a}. Here we report that mice deficient for all three open reading frames (*Cdkn2ab*^{−/−}) are more tumour-prone and develop a wider spectrum of tumours than *Cdkn2a* mutant mice, with a preponderance of skin tumours and soft tissue sarcomas (for example, mesothelioma) frequently composed of mixed cell types and often showing biphasic differentiation. *Cdkn2ab*^{−/−} mouse embryonic fibroblasts (MEFs) are substantially more sensitive to oncogenic transformation than *Cdkn2a* mutant MEFs. Under conditions of stress, p15^{Ink4b} protein levels are significantly elevated in MEFs deficient for p16^{Ink4a}. Our data indicate that p15^{Ink4b} can fulfil a critical backup function for p16^{Ink4a} and provide an explanation for the frequent loss of the complete *CDKN2b–CDKN2a* locus in human tumours.

Both p15^{Ink4b} and p16^{Ink4a} are able to induce cell cycle arrest in G1 by inhibiting cyclin-dependent kinases CDK4 and CDK6 to inactivate the retinoblastoma (RB1) family of tumour suppressor proteins². The unrelated p14^{ARF} protein acts primarily by way of MDM2 to activate the key checkpoint protein TRP53, thereby inducing either cell cycle arrest (both in G1 and in G2) or apoptosis³. The entire locus is frequently lost or mutated in a wide range of primary tumours^{4–8}. Mutational analyses and studies with mouse models indicate that both p16^{Ink4a} and p14^{ARF} function as tumour suppressors^{8–13}. For p15^{Ink4b}, the situation is less clear, because no point mutations affecting its function and only a few cases of epigenetic silencing have been found^{18,14}. Moreover, mice deficient for p15^{Ink4b} show only a very subtle tumour predisposition¹⁵. The frequent loss of cyclin-dependent inhibitor locus 2 (*CDKN2b*) in human tumours could be explained by its juxtaposition to *CDKN2a* leading to frequent co-deletion, although combined loss of p16^{Ink4a} and p15^{Ink4b} is associated with a significant reduction in survival of patients with soft tissue sarcomas⁷.

To study the effect of additional p15^{Ink4b} loss on tumorigenesis, we have generated two new mouse strains, one in which both p15^{Ink4b} and p16^{Ink4a} are inactivated (which we call *Ink4ab*^{−/−}), and one deficient for p15^{Ink4b}, p16^{Ink4a} and p19^{Arf} (which we call *Cdkn2ab*^{−/−}). The *Ink4ab*^{−/−} and *Cdkn2ab*^{−/−} alleles are schematically depicted in Fig. 1a. For more details see Methods and Supplementary Fig. 1.

Within seven months, all *Cdkn2ab*^{−/−} animals develop one or more tumours (Fig. 1b). The predominant lesions are skin tumours, including squamous cell carcinoma, basal cell carcinoma and trichoepithelioma/hair-germ-cell tumours, and various soft tissue sarcomas (Fig. 1c and Fig. 2). Both types of lesions are typically composed of different cell types, possibly reflecting oncogenic transformation of an early progenitor, and frequently show biphasic morphology. For a detailed description and histological characterization of both types of lesions, see Supplementary Data. Several animals developed lymphoma or leukaemia. In addition, a few cases of (adeno)carcinoma of the intestinal tract and one case of spontaneous melanoma (onset

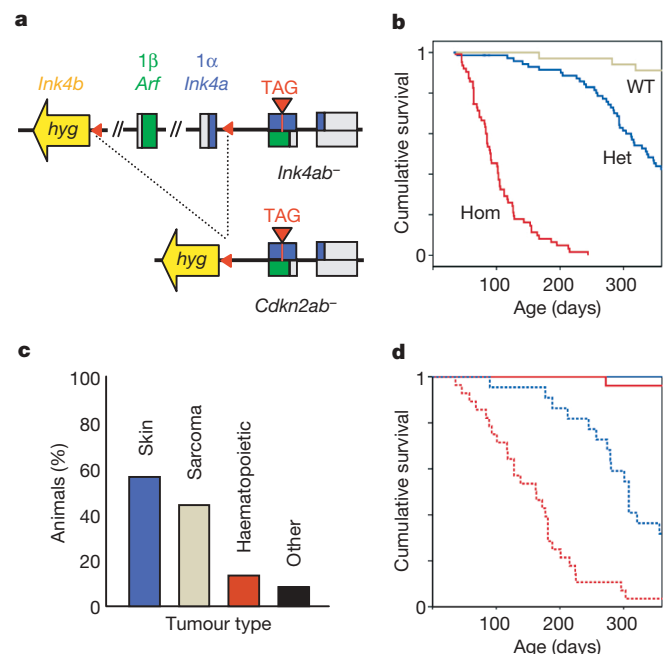


Figure 1 | Generation of *Ink4ab*^{−/−} and *Cdkn2ab*^{−/−} mice. **a**, Schematic representation of the *Ink4ab*^{−/−} and *Cdkn2ab*^{−/−} mutant alleles. hyg, hygromycin cassette; red triangle symbol, loxP site; TAG, p16^{Ink4a}-specific point mutation. **b**, Kaplan–Meier cumulative survival curves of *Cdkn2ab*^{−/−} (homozygous, Hom; n = 55), *Cdkn2ab*^{+/−} (heterozygous, Het; n = 57) and *Cdkn2ab*^{+/+} (wild type, WT; n = 27) littermates. *P* < 0.001 for every pair. **c**, Tumour spectrum of *Cdkn2ab*^{−/−} mice. **d**, Kaplan–Meier cumulative survival curves of DMBA-treated mice (dashed lines represent +DMBA). Blue, *Ink4a*^{−/−} (n = 15 treated, n = 23 untreated); red, *Ink4ab*^{−/−} (n = 27 treated, n = 27 untreated). *P* < 0.001 for treated versus untreated animals, and for treated *Ink4ab*^{−/−} versus treated *Ink4a*^{−/−} animals.

¹Division of Molecular Genetics and Centre for Biomedical Genetics, ²Department of Experimental Animal Pathology, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands. †Present address: GSF, Institute for Stem Cell Research, D-85764 Munich, Germany.

*These authors contributed equally to this work.

at three weeks of age) were observed. Except for skin tumours, *Cdkn2ab*^{+/-} heterozygous animals develop similar tumours to homozygous animals although they have a much longer latency (Fig. 1b). *Ink4ab*^{-/-} animals that are p19^{Arf}-proficient do not show a significant tumour predisposition within the first year (Fig. 1d). However, application of a single dose of dimethylbenzanthrene (DMBA) to newborn *Ink4ab*^{-/-} pups induces the same spectrum of tumours with the same histological features as observed for *Cdkn2ab*^{-/-} mice, including skin tumours. Notably, DMBA-induced tumours appear much earlier and also show a much broader diversity in *Ink4ab*^{-/-} mice than in mice deficient in p16^{Ink4a} (which we call *Ink4a*^{-/-} mice), indicating that p15^{Ink4b} has a critical tumour suppressor activity (Fig. 1d).

To analyse the effects of different *Cdkn2b*–*Cdkn2a* mutations on cell behaviour, MEFs from various strains were characterized using a 3T3 protocol. For a summary of the different lines used throughout this study, see Supplementary Fig. 2. As expected, p19^{Arf}-deficient MEFs grow faster than p19^{Arf}-proficient MEFs. Combined loss of p15^{Ink4b} and p16^{Ink4a}, in both p19^{Arf}-proficient and -deficient cells, confers a

slightly increased growth potential and a shorter cell cycle re-entry latency compared to loss of either p15^{Ink4b} or p16^{Ink4a} alone (Fig. 3a,b).

Soft agar assays showed that none of the cell lines were capable of anchorage-independent growth. After retroviral transduction of oncogenic RAS (*HRAS*(G12V)), only p19^{Arf}-deficient MEFs formed colonies in soft agar (Fig. 4a). Notably, *HRAS*(G12V) transformation of *Cdkn2ab*^{-/-} cells resulted in about tenfold more colonies than transformation of either *Arf*^{-/-} or *Ink4a/Arf*^{-/-} cells. Moreover, both under normal culture conditions and in soft agar, the *HRAS*(G12V)-transformed *Cdkn2ab*^{-/-} cells grew much faster than either *Arf*^{-/-} or *Ink4a/Arf*^{-/-} cells, with the first soft agar colonies apparent after three days. This striking difference cannot be explained by large variations in expression levels of RAS, Cdk6 or phosphorylated (P)-Erk (Supplementary Fig. 5). Transduction of *c-MYC* also increased colony growth in soft agar, although the colonies were smaller than the *HRAS*(G12V)-transformed colonies. In contrast, transformation by both *c-MYC* and *HRAS*(G12V) resulted in roughly equal numbers of soft agar colonies from *Arf*^{-/-}, *Ink4a/Arf*^{-/-} and *Cdkn2ab*^{-/-} cells. These *in vitro* assays suggest that, in *Ink4a/Arf*^{-/-} MEFs, p15^{Ink4b} suppresses oncogenic transformation.

To determine how p15^{Ink4b} and p16^{Ink4a} expression is correlated, we analysed the expression of p15^{Ink4b} and p16^{Ink4a} under different conditions. In proliferating cells, p16^{Ink4a} expression is initially very low and increases in subsequent passages. Expression of p15^{Ink4b} is already significant at passage two, and remains stable throughout the culture process (Fig. 3e). In contrast, on oncogene-induced expression of p16^{Ink4a}, p15^{Ink4b} expression is lost (Fig. 4d). The same correlation between p16^{Ink4a} and p15^{Ink4b} expression levels was observed in super-confluent cells (Supplementary Fig. 4a).

Whereas p16^{Ink4a} transcription was induced on consecutive passaging or introduction of *HRAS*(G12V) (Fig. 3d, 4c and Supplementary Fig. 4b), levels of messenger RNA encoding p15^{Ink4b} remained stable under these conditions (Figs 3c, d, 4c and Supplementary Fig. 4b). Thus, p16^{Ink4a} protein levels reflect RNA levels, whereas p15^{Ink4b} protein levels do not. Notably, treatment of *HRAS*(G12V)-transformed cells with the proteasome inhibitor MG132 causes a 2.6-fold increase in p15^{Ink4b} protein levels in the presence of p16^{Ink4a}, indicating that p15^{Ink4b} is subject to proteasomal degradation. In contrast, MG132 treatment did not increase p15^{Ink4b} protein levels in p16^{Ink4a}-deficient cells (Fig. 4e). Small variations in *Cdkn2d* (which encodes p19^{Ink4d}) mRNA and protein levels were observed, but these did not correlate with either genotype or passage number. However, in the absence of both p16^{Ink4a} and p15^{Ink4b}, slightly higher levels of *Cdkn2c* (which encodes p18^{Ink4c}) mRNA and protein are found (Supplementary Fig. 4c, d).

Binding of p15^{Ink4b} to Cdk4 was increased in the absence of p16^{Ink4a}, both in normal cells (Fig. 3f) and in *HRAS*(G12V)-transformed cells (Fig. 4f). Binding of p18^{Ink4c} to Cdk4 remained unchanged in the different lines and under the various conditions. This shows that p15^{Ink4b} is able to functionally substitute for p16^{Ink4a}, and suggests a unique relationship, at least in MEFs, between p16^{Ink4a} and p15^{Ink4b}.

Because RAS transformation is dependent on Cdk4 (ref. 16), we speculated that co-expression of an Ink4-insensitive CDK4 mutant (CDK4(R24C)) would eliminate the differences observed between the p19^{Arf}-deficient strains (Fig. 4b). Introduction of both *CDK4*(R24C) and *HRAS*(G12V) into MEFs increased the number of *Ink4a/Arf*^{-/-} and *Arf*^{-/-} colonies more than 7.5-fold, but had hardly any effect on the number of *Cdkn2ab*^{-/-} colonies. We also noted that, in soft agar assays, the *Ink4a/Arf*^{-/-} MEFs were less sensitive to transformation than *Arf*^{-/-} MEFs. This could reflect differences in genetic background and/or residual p19^{Arf} activity from the remnant exon 1β in the *Ink4a/Arf*^{-/-} cells that becomes apparent under the experimental conditions used (for example, high levels of oncogenic RAS)¹⁷. Also, we cannot exclude that p15^{Ink4b} has functions that are not shared with p16^{Ink4a}. However, our data indicate that both p16^{Ink4a} and

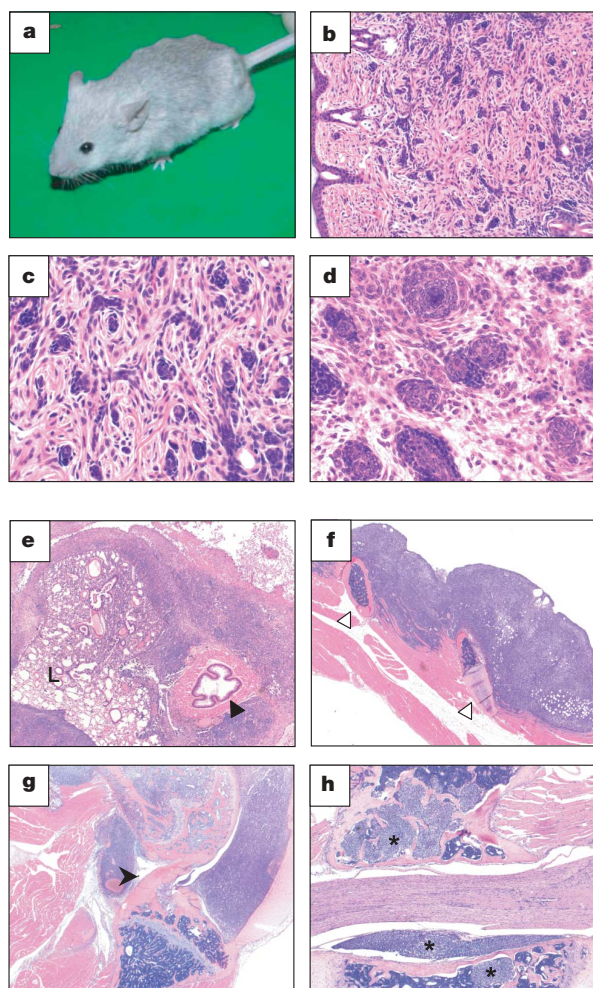


Figure 2 | *Cdkn2ab*^{-/-} mice develop skin tumours and sarcomatous lesions. **a**, A *Cdkn2ab*^{-/-} mouse with multiple skin lesions. **b**, H&E staining (×20) of a well-differentiated squamous cell carcinoma. **c**, A detailed image (H&E, ×40) of the lesion shown in **b**, showing basaloid and hair-germ-cell-like differentiation. **d**, A detailed image (H&E, ×40) showing extensive hair germ-like differentiation. **e**, Mesothelioma-like lesion in the lung, pleura and mediastinum (H&E, ×2.5). L, lung; the black arrowhead indicates the oesophagus. **f**, Mesothelioma-like lesion in the thoracic wall (H&E, ×2.5). The white arrowheads indicate the ribs. **g**, Synovial sarcoma-like lesion surrounding the knee joint (H&E, ×2.5). The black arrowhead indicates the cross ligament. **h**, Lesion surrounding the vertebral column (H&E, ×5). The asterisks represent tumours.

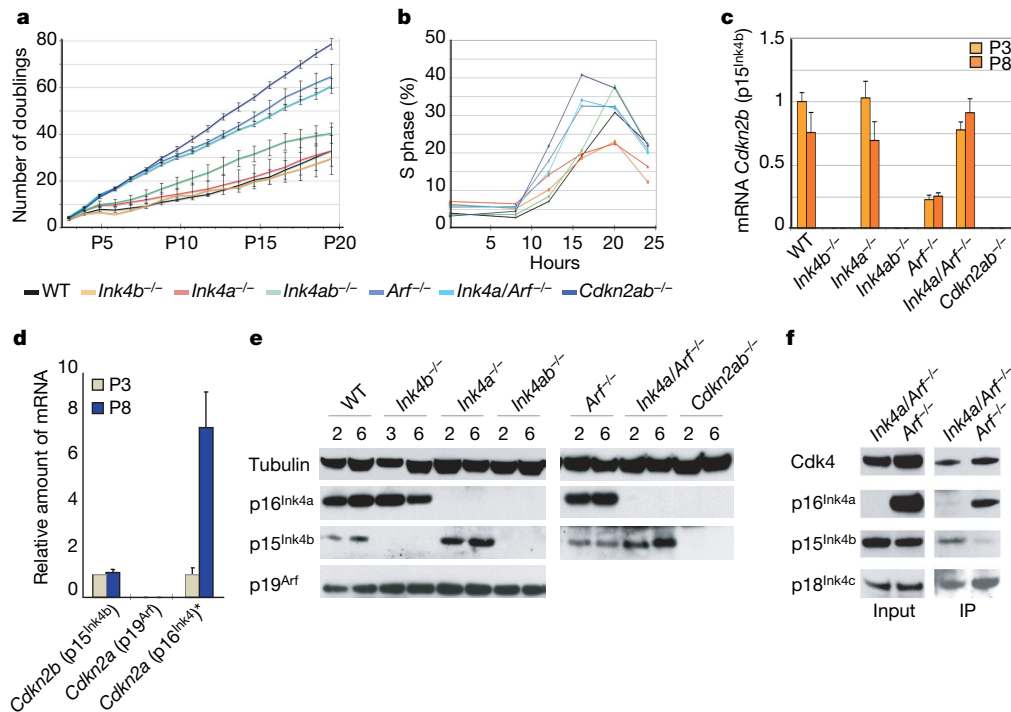


Figure 3 | MEF analysis. **a**, Proliferation curves (\pm s.d., $n \geq 3$). **P**, passage number. **b**, Cell cycle re-entry curves. Horizontal axis shows number of hours after the addition of serum. **c**, Relative mRNA levels (Taqman; \pm s.d., $n \geq 4$). Levels in wild-type (WT) cells at P3 were set at 1. **d**, Relative mRNA levels in p19^{Arf}-deficient cells (Taqman; \pm s.d., $n \geq 4$). Levels at P3 were set at 1. **e**, Relative protein levels in proliferating cells. Passage numbers are indicated. **f**, Cdk4 binding. IP, anti-Cdk4-precipitated *Ink4a/Arf*^{-/-} and *Arf*^{-/-} extracts.

p15^{Ink4b} suppress *HRAS*(G12V) oncogenic transformation primarily by means of inhibition of Cdk4 signalling. Taken together, our results show that loss of the complete *Cdkn2b*–*Cdkn2a* locus confers a very high tumour predisposition. Mice in which p19^{Arf} is knocked out are primarily predisposed to fibrosarcoma and lymphoma after a relatively long latency^{10,11}. Additional loss of p16^{Ink4a} results in a closely resembling tumour phenotype^{9,18}. Here we show that

1. Asterisk, levels of mRNA encoding p16^{Ink4a} were measured using a probe measuring both p16^{Ink4a} and p19^{Arf}. **e**, Relative protein levels in proliferating cells. Passage numbers are indicated. **f**, Cdk4 binding. IP, anti-Cdk4-precipitated *Ink4a/Arf*^{-/-} and *Arf*^{-/-} extracts.

loss of the second inhibitor of Cdk4 and 6 encoded by this locus results in a pronounced predisposition to the development of a wide range of highly malignant tumours not seen in either *Arf*^{-/-} or *Ink4a/Arf*^{-/-} mice. Notably, after DMBA treatment, *Ink4ab*^{-/-} mice show a much higher tumour predisposition than either *Ink4a*^{-/-} (Fig. 1d) or *Ink4b*^{-/-} mice¹⁵, indicating that the combined loss of p15^{Ink4b} and p16^{Ink4a} augments tumour predisposition independent of the p19^{Arf}

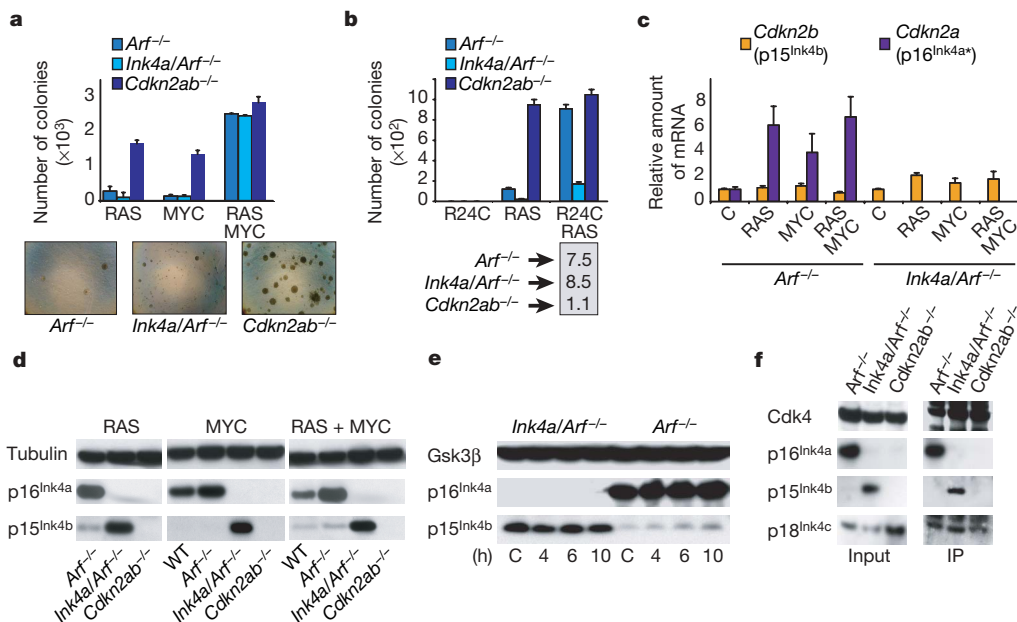


Figure 4 | Analysis of oncogenically transformed MEFs. **a**, Soft agar analysis of *c-MYC* (MYC)- and *HRAS*(G12V) (RAS)-transformed cells. Images show *HRAS*(G12V)-transformed cells after ten days of culture (\pm s.d., $n = 3$). **b**, Soft agar analysis of *Cdk4*(R24C)- (called R24C in image) and *HRAS*(G12V) (RAS)-transformed cells. The fold increase in the number of colonies in doubly transformed cells as compared to in the *HRAS*(G12V)

single transformation is indicated below the graph (\pm s.d., $n = 2$). **c**, Relative mRNA levels (Taqman; \pm s.d., $n \geq 4$). Levels in mock-transformed (C) cells were set at 1. Asterisk, the levels of p16^{Ink4a} were obtained using a probe measuring both p16^{Ink4a} and p19^{Arf}. **d**, Relative protein expression levels. **e**, Relative protein levels after MG132 treatment. The length of treatment is indicated (hours). **f**, Cdk4 binding in *HRAS*(G12V)-transformed cells.

status. The strong effect of co-deletion of $p15^{\text{Ink4b}}$ and $p16^{\text{Ink4a}}$ could be explained by attributing distinct complementing functions to these two proteins. We consider this to be unlikely for the following reasons. First, hardly any *INK4b* mutations not affecting *INK4a* have been reported. Second, $p15^{\text{Ink4b}}$ knockout mice have a weak tumour phenotype and $p15^{\text{Ink4b}}$ -deficient MEFs show only a slight increase in sensitivity to transformation by oncogenic *RAS* (ref. 15). Third, our *in vitro* data show that $p15^{\text{Ink4b}}$ protein is degraded when $p16^{\text{Ink4a}}$ is expressed. Last, binding of $p15^{\text{Ink4b}}$ to Cdk4 is increased in the absence of $p16^{\text{Ink4a}}$. Therefore, our results are best explained by a model in which $p15^{\text{Ink4b}}$ fulfils a 'backup' role for $p16^{\text{Ink4a}}$.

It has been suggested that $p16^{\text{Ink4a}}$ is the most effective inhibitor of CDKs 4 and 6, because only $p16^{\text{Ink4a}}$ can form stable, binary complexes with both CDK4 and CDK6 (ref. 19). However, $p16^{\text{Ink4a}}$ is expressed at very low levels under normal physiological conditions. Under conditions of stress (including oncogenic stress, ageing and growth inhibition), the expression of $p16^{\text{Ink4a}}$ increases. Should $p15^{\text{Ink4b}}$ be expressed, $p16^{\text{Ink4a}}$ will efficiently compete for Cdk4 (and probably Cdk6) binding, resulting in the displacement and degradation of the $p15^{\text{Ink4b}}$ protein. Many cells express $p15^{\text{Ink4b}}$ either constitutively or on induction by factors such as transforming growth factor- β . On loss of $p16^{\text{Ink4a}}$, Cdk4 inhibition is maintained by $p15^{\text{Ink4b}}$. In $p19^{\text{Arf}}$ -deficient MEFs, the loss of both $p16^{\text{Ink4a}}$ and $p15^{\text{Ink4b}}$ only marginally affects their growth properties, but greatly enhances their propensity to become transformed by oncogenic *RAS*. This might relate to the earlier reported ability of $p16^{\text{Ink4a}}$ to inhibit integrin-mediated cell spreading²⁰, which could explain the early onset and aggressive growth of tumours in *Cdkn2ab*^{-/-} mice. According to our model, efficient malignant transformation would require the inactivation of both $p16^{\text{Ink4a}}$ and $p15^{\text{Ink4b}}$, thus providing a rationale as to why the entire *CDKN2a-CDKN2b* locus is so frequently deleted. Although the other family members $p18^{\text{INK4c}}$ and $p19^{\text{INK4d}}$ could have a role similar to that of $p15^{\text{Ink4b}}$, there are no reports showing the combined loss of $p16^{\text{Ink4a}}$ and either of these proteins. The duplication of the gene that encodes $p15^{\text{Ink4b}}$, which is considered the ancestral Ink4 inhibitor in this locus, might have evolved in mammals to serve a dual function¹. As a result, $p15^{\text{Ink4b}}$ and $p16^{\text{Ink4a}}$ now respond to different signalling pathways, although acting mostly on the same target proteins. At the same time, $p15^{\text{Ink4b}}$ serves as a backup of $p16^{\text{Ink4a}}$, thereby ensuring a more robust tumour suppressive capacity of this important locus. Even so, this remains a hazardous situation, because a single deletion could result in the elimination of three critical tumour suppressor activities.

METHODS SUMMARY

The *Ink4ab*^{-/-} strain was obtained through *Ink4b* targeting in *Ink4a*⁺ ES cells (in which there is a point mutation in exon 2 of *Cdkn2a* that is silent for the $p19^{\text{Arf}}$ reading frame but results in a stop codon in the $p16^{\text{Ink4a}}$ reading frame)¹³. The *Cdkn2ab*^{-/-} strain was obtained through subsequent crossing of the *Ink4ab*^{-/-} strain with the actin-Cre deleter strain. For tumorigenic studies, mice were killed when seriously ill or when visible tumours reached a diameter of 1 cm. DMBA treatment consisted of the application of a single dose of DMBA (50 μ l of a solution of 0.5% DMBA in acetone) to the dorsal skin of newborn pups (0–3 days old).

Tissues were fixed in either 4% formaline or acidified formalin (ethanol/acetic acid/formaldehyde/saline at 40:5:10:45 v/v)²¹. Haematoxylin and eosin (H&E) stainings were performed according to standard procedures. A full list of all the antibodies used can be found in the Methods.

MEFs from E13.5 embryos were isolated and cultivated according to standard protocols²². Cell cycle re-entry was determined after 72 h of serum starvation (0.5% FCS). Oncogenic transformations were performed according to standard protocols using the following viruses: pLZRS-ires-gfp-hygro, pLZRS-cMYC-ires-gfp, pLZRS-CDK4(R24C)-ires-gfp, pBabePuro and pBabePuro-HRAS(G12V).

The immunoprecipitations were performed essentially as described in ref. 23, and all immunocomplexes were visualized by enhanced chemiluminescence (Pierce 34080).

RNA expression analysis was performed using both Taqman and SYBR Green quantitative PCR methods.

FACS analysis was performed according to standard protocols using a FACSCalibur flowcytometer (Becton Dickinson) and CellQuest software (Becton Dickinson) for data analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 May; accepted 9 July 2007.

- Gil, J. & Peters, G. Regulation of the *INK4b-ARF-INK4a* tumour suppressor locus: all for one or one for all. *Nature Rev. Mol. Cell Biol.* **7**, 667–677 (2006).
- Ortega, S., Malumbres, M. & Barbacid, M. Cyclin D-dependent kinases, *INK4* inhibitors and cancer. *Biochim. Biophys. Acta* **1602**, 73–87 (2002).
- Pomerantz, J. et al. The *Ink4a* tumor suppressor gene product, $p19^{\text{Arf}}$, interacts with MDM2 and neutralizes MDM2's inhibition of p53. *Cell* **92**, 713–723 (1998).
- Drexler, H. G. Review of alterations of the cyclin-dependent kinase inhibitor *INK4* family genes *p15*, *p16*, *p18* and *p19* in human leukemia-lymphoma cells. *Leukemia* **12**, 845–859 (1998).
- Krug, U., Ganser, A. & Koeffler, H. P. Tumor suppressor genes in normal and malignant hematopoiesis. *Oncogene* **21**, 3475–3495 (2002).
- Murao, K., Kubo, Y., Ohtani, N., Hara, E. & Arase, S. Epigenetic abnormalities in cutaneous squamous cell carcinomas: frequent inactivation of the *RB1/p16* and *p53* pathways. *Br. J. Dermatol.* **155**, 999–1005 (2006).
- Orlow, I. et al. Alterations of *INK4A* and *INK4B* genes in adult soft tissue sarcomas: effect on survival. *J. Natl. Cancer Inst.* **91**, 73–79 (1999).
- Ruas, M. & Peters, G. The $p16^{\text{Ink4a}}$ /CDKN2A tumor suppressor and its relatives. *Biochim. Biophys. Acta* **1378**, F115–F177 (1998).
- Serrano, M. et al. Role of the *INK4a* locus in tumor suppression and cell mortality. *Cell* **85**, 27–37 (1996).
- Kamijo, T. et al. Tumor suppression at the mouse *INK4a* locus mediated by the alternative reading frame product $p19^{\text{ARF}}$. *Cell* **91**, 649–659 (1997).
- Kamijo, T., Bodner, S., van de Kamp, E., Randle, D. H. & Sherr, C. J. Tumor spectrum in *ARF*-deficient mice. *Cancer Res.* **59**, 2217–2222 (1999).
- Sharpless, N. E. et al. Loss of $p16^{\text{Ink4a}}$ with retention of $p19^{\text{Arf}}$ predisposes mice to tumorigenesis. *Nature* **413**, 86–91 (2001).
- Krimpenfort, P., Quon, K. C., Mooi, W. J., Loonstra, A. & Berns, A. Loss of $p16^{\text{Ink4a}}$ confers susceptibility to metastatic melanoma in mice. *Nature* **413**, 83–86 (2001).
- Kim, W. Y. & Sharpless, N. E. The regulation of *INK4/ARF* in cancer and aging. *Cell* **127**, 265–275 (2006).
- Latres, E. et al. Limited overlapping roles of $P15^{\text{Ink4b}}$ and $P18^{\text{INK4c}}$ cell cycle inhibitors in proliferation and tumorigenesis. *EMBO J.* **19**, 3496–3506 (2000).
- Zou, X. et al. Cdk4 disruption renders primary mouse cells resistant to oncogenic transformation, leading to *Arf*/p53-independent senescence. *Genes Dev.* **16**, 2923–2934 (2002).
- Pantoja, C., Palmero, I. & Serrano, M. Identification of the gene immediately downstream of the murine *INK4a/ARF* locus. *Exp. Gerontol.* **36**, 1289–1302 (2001).
- Sharpless, N. E., Ramsey, M. R., Balasubramanian, P., Castrillon, D. H. & DePinho, R. A. The differential impact of $p16^{\text{Ink4a}}$ or $p19^{\text{ARF}}$ deficiency on cell growth and tumorigenesis. *Oncogene* **23**, 379–385 (2004).
- Parry, D., Mahony, D., Wills, K. & Lees, E. Cyclin D-CDK subunit arrangement is dependent on the availability of competing *INK4* and *p21* class inhibitors. *Mol. Cell. Biol.* **19**, 1775–1783 (1999).
- Fahraeus, R. & Lane, D. P. The $p16^{\text{Ink4a}}$ tumour suppressor protein inhibits $\alpha_v\beta_3$ integrin-mediated cell spreading on vitronectin by blocking PKC-dependent localization of $\alpha_v\beta_3$ to focal contacts. *EMBO J.* **18**, 2106–2118 (1999).
- Harrison, P. T. An ethanol-acetic acid-formol saline fixative for routine use with special application to the fixation of non-perfused rat lung. *Lab. Anim.* **18**, 325–331 (1984).
- Deng, C., Zhang, P., Harper, J. W., Elledge, S. J. & Leder, P. Mice lacking $p21^{\text{Cip1/Waf1}}$ undergo normal development, but are defective in G1 checkpoint control. *Cell* **82**, 675–684 (1995).
- Bagui, T. K., Mohapatra, S., Haura, E. & Pledger, W. J. $P27^{\text{Kip1}}$ and $p21^{\text{Cip1}}$ are not required for the formation of active D cyclin-cdk4 complexes. *Mol. Cell. Biol.* **23**, 7285–7290 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. B. Ali for assistance in generating the mice, the staff of the NKI animal facility for providing animal care, the staff of the histology department for the processing of tissues, J. de Ridder for help with statistical analysis, L. Vredevel and C. Michaloglou for help with soft agar assays, and R. van Amerongen and the group of D. Peepers for discussions. This work was supported by the Dutch Cancer Society.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.B. (a.berns@nki.nl).

METHODS

Mice. For generation of *Ink4ab*^{-/-} and *Cdkn2ab*^{-/-} mouse strains, see Supplementary Fig. 1. The *Ink4a*⁺ ES cell clone¹³, containing a *PGK-neo* selection cassette flanked by *loxP* sites 1 kbp upstream of exon 2, was used as starting material. In the p15^{Ink4b} targeting vector, *Cdkn2b* exons 1 and 2 were replaced by a *PGK-hyg* selection cassette (orientation indicated in Supplementary Fig. 1a) with a 5' *loxP* site in the same orientation as the *loxP* sites flanking the *PGK-neo* cassette upstream of exon 2 of *Cdkn2a*. Correctly targeted clones were selected by Southern blotting using an *XbaI/BamHI* probe 9 kb downstream of exon 2 of *Cdkn2b*. Subsequently, each of these clones was transfected transiently with a Cre recombinase plasmid to identify the clones in which the genes encoding p15^{Ink4b} and p16^{Ink4a} were disrupted on the same allele. When the modifications are on different alleles, Cre-mediated deletion will result in the removal of only the *PGK-neo* cassette. In contrast, in clones containing a doubly targeted *Cdkn2b-Cdkn2a* allele, Cre activity can result in three distinct deletion products depending on the specific set of *loxP* sites involved. One parental clone gave separate subclones in which all three of these Cre-mediated deletions were detected. From this parental clone, subclones in which only the *PGK-neo* cassette had been deleted (leaving one *loxP* site 3' of *Cdkn2b* exon 2 and a second *loxP* site 5' of *Cdkn2a* exon 2) were used for blastocyst injection. F₁ offspring were either used to derive a (*Ink4ab*^{-/-}) colony or were crossed with actin-Cre mice²⁴ to generate germline deletion of the entire *Cdkn2a-Cdkn2b* locus (*Cdkn2ab*^{-/-} colony). Both colonies were crossed to an FVB/N background.

For an overview of all strains used throughout this study, please refer to Supplementary Fig. 2. *Ink4a*^{-/-} and *Ink4a/Arf*^{-/-} were published previously as *Ink4a*⁺ and *Ink4a*^{Δ2,3}, respectively¹³. *Ink4b*^{-/-} was generated in our laboratory (unpublished) and carries the same mutation as the published *Ink4b* knockout strain¹⁵. The *Arf*^{-/-} strain was obtained from the laboratory of C. Sherr¹⁰. The second *Cdkn2b-Cdkn2a* knockout strain briefly referred to in the Supplementary Data is depicted as *Cdkn2ab*^{-/-} (A) in the Supplementary Fig. 2. This strain gave similar results as the *Cdkn2ab*^{-/-} (B) strain reported in this study.

Histological analysis. Primary antibodies were obtained from the following sources: E-cadherin, Transduction Laboratories 610182; involucrin, Babco PRB-140C; K1, Babco PRB-165P; K8, Developmental Studies Hybridoma Bank Troma-1; K10, Sigma C7284; K14, Babco PRB-155P; loricrin, Babco PRB-145P; and vimentin, Research Diagnostics RDI-PROGP-53. The mesothelin antibody was a gift from T. Bera. Antibody staining was revealed using either diaminobenzidine or 3-amino-9-ethylcarbazole chromogen. Sections were counterstained with haematoxylin. Images were captured using a Zeiss Axioskop2 Plus microscope (Carl Zeiss Microscopy) equipped with a Zeiss AxioCam HRC digital camera, and processed using AxioVision 4 software (Carl Zeiss Vision).

Cell culture. Cells were grown at 37 °C, 5% CO₂, in DMEM supplemented with 10% FCS, 100 units per ml penicillin, 100 μg ml⁻¹ streptomycin and 0.1 mM β-mercapto-ethanol. Cells were serially passaged twice a week according to the 3T3 protocol. Plates were scraped and pellets were stored at -80 °C for future protein and RNA analysis (see below).

For cell cycle re-entry experiments, cells (passage 2) were serum-starved (0.5% FCS) for 72 h. One hour before analysis, BrdU was added to the medium (10 μM final concentration). Cells were collected by trypsinisation, washed, fixed in 70% ethanol and stored at 4 °C until further analysis.

For soft agar growth analysis, cells (24,000 cells per well in a 6-well plate) were resuspended in 0.4% soft agar and plated on a bottom layer of 1% agar. The number of colonies growing in soft agar after two weeks of culture was quantified using ImagePro Plus 5.1.2 software (MediaCybernetics).

For oncogenic transformation, passage two MEFs were infected overnight with pLZRS-ires-gfp-hygro control virus, with pLZRS-cMYC-ires-gfp virus expressing human c-MYC or with pLZRS-CDK4(R24C)-ires-gfp expressing the non-INK4 binding mutant of CDK4. After recovery during the day, the cells were infected overnight with either pBabePuro control virus or pBabePuro-HRAS(G12V) virus expressing human HRAS(G12V). Puromycin selection (4 μg ml⁻¹) was started 24 h after the second infection and maintained for one week. Viruses were provided by D. Peepers and R. van Amerongen.

To inhibit proteasome-mediated protein degradation, HRAS(G12V)-transformed *Arf*^{-/-} and *Ink4a/Arf*^{-/-} cells were treated with either 20 μM MG132 or ethanol for 4, 6 or 10 h.

Protein expression analysis. MEF pellets were lysed in RIPA buffer and the lysates were cleared by centrifugation at 20,800g for 10 min. Proteins were separated on 4–12% NuPAGE Novex Bis-Tris gels (Invitrogen), transferred to nitrocellulose (Schleicher and Schuell) or Immobilon (Millipore) membrane and probed with antibodies. Immuno-complexes were visualized by enhanced chemiluminescence (Pierce 34080) using goat anti-mouse (AM13404, Biosource) or goat anti-rabbit (ALI0404, Biosource) antibodies coupled to horseradish peroxidase.

Cells were harvested by scraping and were incubated in lysis buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM EDTA, 0.5% NP-40, 10% glycerol, 0.1 mM Na₃VO₄, 0.5 mM NaF, 0.1 mM PMSF, 2.5 μg of leupeptin per ml and 1 mM DTT) for 30 min on ice. After overnight incubation with anti-Cdk4 at 4 °C, the immune complexes were recovered with protein G-agarose beads (1 to 2 h, 4 °C) and washed twice with lysis buffer. The immune complexes were resolved on a 10% SDS-PAGE gel and electrophoretically transferred to nitrocellulose membranes. Membranes were blocked in PBS containing 0.1% Tween 20 and 5% instant milk, and were incubated with antibody for 2 h at room temperature (18–21 °C) in PBS containing 0.1% Tween 20. Proteins recognized by the antibody were detected by enhanced chemiluminescence using a horseradish peroxidase-coupled secondary antibody (Pierce).

The primary antibodies used were: p16^{Ink4a}, M-156 (Santa Cruz); p15^{Ink4b}, K-18 (Santa Cruz); p18^{Ink4c}, M-168 (Santa Cruz); p19^{Ink4d}, M-167 (Santa Cruz); p19^{Arf}, R562 (Abcam); α-tubulin, B512 (Sigma); RAS, 3965 (Cell Signaling); P-Erk, 9182 (Cell Signaling); Cdk4, sc-260 (Santa Cruz); and Cdk6, 3136 (Cell Signaling).

RNA expression analysis. RNA was isolated from frozen MEF pellets using RNeasy columns (Qiagen). For each sample, 1 μg mRNA was reverse transcribed using random hexamers (Roche, first Strand cDNA synthesis kit for RT-PCR). Samples were analysed at least in triplicate on a minimum of two different plates. Reactions were performed using the ABI Prism 7000 sequence detection system.

Taqman quantitative PCR analysis. Reactions were performed according to manufacturer's instructions using Taqman Universal PCR Master Mix (Roche) and the equivalent of 10 ng of RNA per reaction. The following probe/primer mixes were used: *Cdkn2b* (p15^{Ink4b}) Mm00483241_m1; *Cdkn2a* (p16^{Ink4a} + p19^{Arf}) Mm00494449_m1; *Cdkn2a* (p19^{Arf}) Mm01257348_m1; *Cdkn2c* (p18^{Ink4c}) Mm01258411_g1; *Cdkn2d* (p19^{Ink4d}) Mm00486943_m1; and Gapdh internal control (Mm99999915_g1).

SYBR Green quantitative PCR analysis. Reactions were performed according to manufacturer's instructions using Power SYBR Green PCR Master Mix (ABI) and the equivalent of 50 ng of RNA per reaction. The following primers were used: p15^{Ink4b}, AGATCCCAACGCCCTGAAC and CCCATCATCATGACCTGGATT; p16^{Ink4a}, CGTACCCCGATTTCAGGTGAT and TTGAGCAGAAGAGCTGCTACGT; p19^{Arf}, GCCGCACCGGAATCCT and TTGAGCAGAAGAGCTGCTACGT; and Hprt internal control, CTGGTGAAAGGACCTCTCG and TGAAGTACTCATTATAGTCAAGGGCA.

FACS analysis. BrdU labelling and propidium iodide staining were performed according to standard protocols. Briefly, ethanol-fixed cells were washed (0.5% Tween-20 in PBS), the nuclei were dissociated with HCl/Triton (5 M HCl, 0.5% Triton) for 20 min at room temperature, after which 10 volumes of 0.1 M Na₂B₄O₇ were added for neutralization. The cells were collected by centrifugation, washed again and stained with anti-BrdU-Alexa488 (1:100, Invitrogen A21303) for 30 min at room temperature in wash buffer containing 1% BSA. After two more wash steps, the cells were resuspended in FACS buffer (PBS containing 1% BSA) containing 10 μg ml⁻¹ propidium iodide. Cells were measured using a FACSCalibur flowcytometer (Becton Dickinson) and analysed using CellQuest software (Becton Dickinson).

24. Lyons, S. K., Meuwissen, R., Krimpenfort, P. & Berns, A. The generation of a conditional reporter that enables bioluminescence imaging of Cre/loxP-dependent tumorigenesis in mice. *Cancer Res.* 63, 7042–7046 (2003).

The effects of molecular noise and size control on variability in the budding yeast cell cycle

Stefano Di Talia^{1,2}, Jan M. Skotheim², James M. Bean^{1†}, Eric D. Siggia² & Frederick R. Cross¹

Molecular noise in gene expression can generate substantial variability in protein concentration¹. However, its effect on the precision of a natural eukaryotic circuit such as the control of cell cycle remains unclear. We use single-cell imaging of fluorescently labelled budding yeast to measure times from division to budding (G1) and from budding to the next division. The variability in G1 decreases with the square root of the ploidy through a 1N/2N/4N ploidy series, consistent with simple stochastic models for molecular noise. Also, increasing the gene dosage of G1 cyclins decreases the variability in G1. A new single-cell reporter for cell protein content allows us to determine the contribution to temporal G1 variability of deterministic size control (that is, smaller cells extending G1). Cell size control contributes significantly to G1 variability in daughter cells but not in mother cells. However, even in daughters, size-independent noise is the largest quantitative contributor to G1 variability. Exit of the transcriptional repressor Whi5 from the nucleus partitions G1 into two temporally uncorrelated and functionally distinct steps. The first step, which depends on the G1 cyclin gene *CLN3*, corresponds to noisy size control that extends G1 in small daughters, but is of negligible duration in mothers. The second step, whose variability decreases with increasing *CLN2* gene dosage, is similar in mothers and daughters. This analysis decomposes the regulatory dynamics of the Start transition into two independent modules, a size sensing module and a timing module, each of which is predominantly controlled by a different G1 cyclin.

In budding yeast (*Saccharomyces cerevisiae*), smaller cells delay the Start transition in late G1, which could imply that cell cycle initiation has a deterministic requirement for a critical size or translation rate^{2–5}. However, in apparent contradiction to deterministic models, the timing of G1 in budding yeast shows substantial variability that is independent of cell size^{6–9}. Here, we analyse the conflicting roles of deterministic cell size control and molecular noise in gene expression^{1,9,10} in generating and controlling cell cycle variability.

We first address the hypothesis that gene expression noise¹ generates significant cell cycle variability. Doubling ploidy doubles the number of all cellular components, but cell volume also doubles, so the average concentrations of cellular components remain constant¹¹. If stochastic variation in the numbers of key molecules causes gene expression noise, then doubling their average number along with the ploidy should reduce this noise (expressed as standard deviation divided by the mean) by about $\sqrt{2}$ (ref. 12). If cell cycle timing is controlled by noisy gene expression, then doubling ploidy should similarly reduce cell cycle variability.

We measured times from cytokinesis to budding (G1) and from budding to cytokinesis in haploids, diploids or tetraploids (mothers and daughters), using time-lapse fluorescence microscopy of strains expressing Myo1 tagged with green fluorescent protein

(Myo1–GFP). Myo1 forms a ring at the new bud neck¹³ (concomitant with initiation of DNA replication²), which disappears at cytokinesis¹³ (Fig. 1). G1 variability is reduced in both mothers and daughters by about the expected factor of $\sqrt{2}$ for each ploidy doubling (Fig. 1, Supplementary Table 8, Supplementary Fig. 7). Thus molecular noise might be an important source of G1 variability. In marked contrast, variability in the time from budding to cytokinesis is nearly unaffected by ploidy (Supplementary Table 8).

The magnitude and sensitivity to ploidy of G1 noise indicate that the noise might be due to small variable numbers of key regulatory molecules. As G1 cyclins control average G1 duration^{14–22} and are expressed at a level of few messenger RNA transcripts per cell²³, we integrated tandem arrays of the G1 cyclin genes *CLN2* or *CLN3* in haploids. Increasing copy numbers of G1 cyclin genes decreases G1 variability (Fig. 1, Supplementary Table 10), which could help to explain the reduction in variability that is caused by the presence of multiple copies of the entire genome (Fig. 1).

This analysis is based solely on timing; however, cell size has long been proposed as a deterministic regulator of the Start transition^{3,11}. To provide a protein-based single cell marker of cell size, we placed the DsRed Red Fluorescent Protein under the control of the promoter of the constitutive, strongly expressed actin (*ACT1*) gene. Assuming that the DsRed transcript accumulates and is translated in parallel with bulk cellular mRNA, then total red fluorescence per cell will reflect total cell protein content. When we quantified total red fluorescence per cell as described²⁴, we found exponential growth in single cells (Fig. 2a; Supplementary Information), as deduced previously from pulse-labelling of size-selected populations²⁵. The single-cell growth rate α is moderately variable, but its average agrees well with the bulk culture growth rate (Fig. 2b, Supplementary Table 12). Total red fluorescence scales linearly with ploidy (Supplementary Table 11) and with geometric estimates of cell size. However, using our methods, DsRed fluorescence is a more reliable indicator of cell size than geometric volume estimation (see Supplementary Information). Total red fluorescence for a colony increases exponentially (Fig. 2c), so changes in the microenvironment do not interfere with these measurements. These results support the use of total red fluorescence from *ACT1pr–DsRed* as a single-cell marker for cell size.

Size control at Start would require smaller cells to prolong G1 for growth, thereby linking birth size and G1 duration. Given exponential growth, the size at budding, M_{bud} , is related to the size at birth M_{birth} , through the amount of time spent in G1 by the simple formula: $M_{\text{bud}} = M_{\text{birth}} e^{\alpha T_{\text{G1}}}$, where α is the growth rate for exponential growth and T_{G1} represents the overall duration of G1. This expression yields: $\alpha T_{\text{G1}} = \ln(M_{\text{bud}}) - \ln(M_{\text{birth}})$. Plotting correlations between αT_{G1} and $\ln(M_{\text{birth}})$ allows us to distinguish between two classical concepts for G1 control: timers and sizers^{26,27}. If G1 duration is under the control of a timer, then αT_{G1} will be independent of cell

¹The Rockefeller University, ²Center for Studies in Physics and Biology, The Rockefeller University, New York, New York 10021, USA. [†]Present address: Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA.

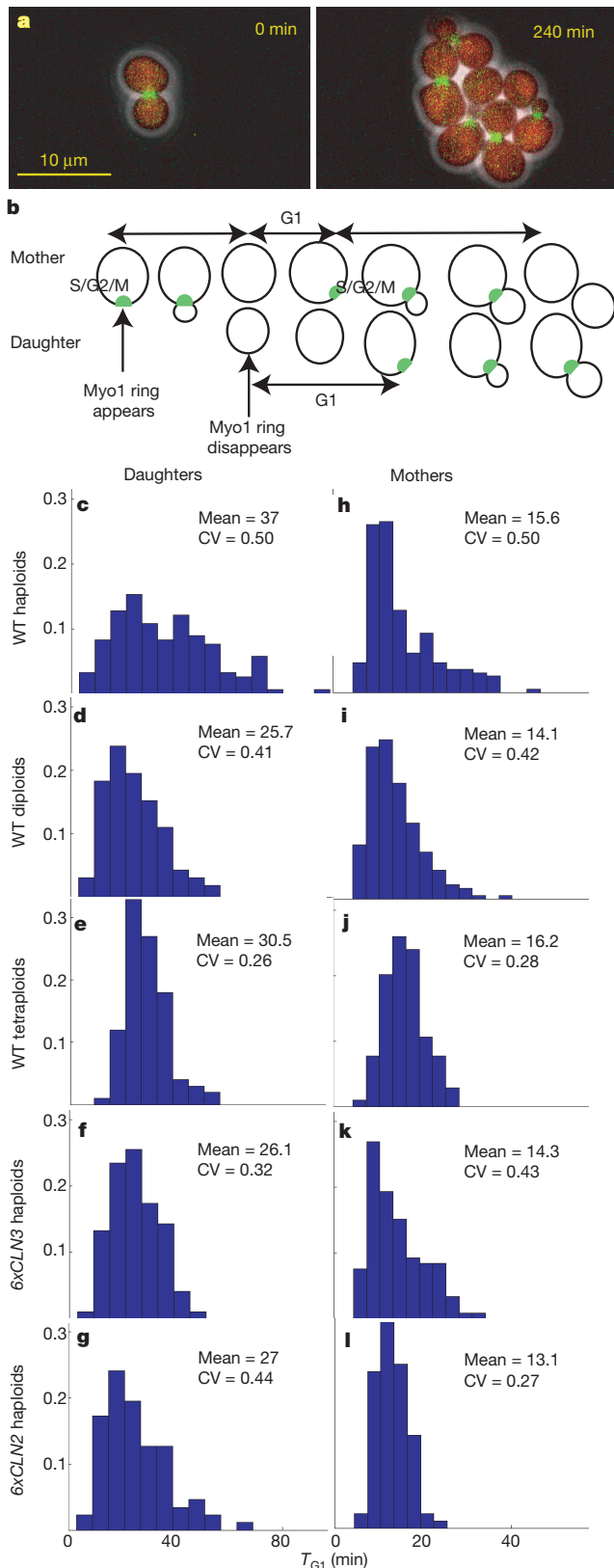


Figure 1 | Noise in G1 duration is reduced by increased ploidy or increased G1 cyclin gene dosage. **a**, Composite phase contrast, Myo1-GFP and *ACT1pr*-DsRed images for haploid cells. **b**, Illustration of measured intervals. **c–l**, Frequency histograms ($n = 87–202$) of the duration of G1 for wild-type (WT) haploid (**c**, **h**), diploid (**d**, **i**) and tetraploid (**e**, **j**), haploid 6xCLN3 (**f**, **k**), and haploid 6xCLN2 (**g**, **l**), daughters (**c**, **d**, **e**, **f**, **g**) and mothers (**h**, **i**, **j**, **k**, **l**). Insets: mean and coefficient of variation (CV: s.d./mean, a standardized noise measure).

size at birth, and the slope of the linear fit of the plot of αT_{G1} against $\ln(M_{\text{birth}})$ will be 0. By contrast, if G1 is controlled by a sizer, all cells will bud at the same size M_{bud} , independent of their size at birth, implying that the slope of the linear fit of αT_{G1} against $\ln(M_{\text{birth}})$ will be -1 (ref. 27).

For the following analysis, rigorous statistical testing of fits is described in the Supplementary Information.

Scaled G1 duration in mother cells is essentially independent of cell size (slope ≈ -0.1), showing that mother G1 is controlled by a timer (Fig. 2d, Supplementary Fig. 9). Daughters, by contrast, show stronger size control (slope ≈ -0.4). Binning the daughter data (Fig. 2e inset) suggested decomposition into two segments, one for small newborn daughters ($<67\%$ of the average budding size), in which an efficient sizer was deduced (slope ≈ -0.7), and a second segment for larger-born daughters, which showed much less dependence on cell size (slope ≈ -0.3 ; Fig. 2e). Statistical confidence in this decomposition was limited by the small number of very small daughters obtained; therefore, we employed the genetic method described in ref. 17 to make unusually small wild-type daughter cells by transient expression of conditional *MET3-CLN2* (see Supplementary Information). Inclusion of these data (Fig. 2f) provided strong statistical support for the two-slope model (linear fit: $P < 0.05$; two-slope fit: $P > 0.7$).

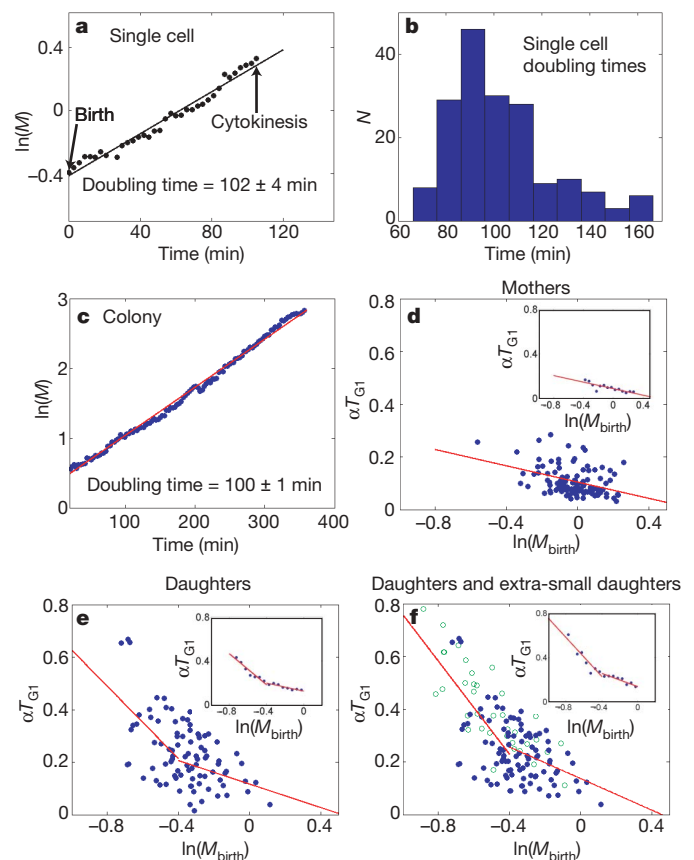


Figure 2 | The correlation between cell size and G1 duration shows that a noisy size control operates in daughters. **a**, Logarithm of total DsRed fluorescence (M) per cell in a single representative cell from birth to cytokinesis. Doubling time is $\ln(2)/\alpha$, where α is the slope of the linear fit. **b**, Haploid cell doubling time distribution. **c**, Total DsRed fluorescence in an entire colony over time. **d**, **e**, Correlation between αT_{G1} (growth-rate-standardized time in G1) and $\ln(M)$ for haploid mothers (**d**) and daughters (**e**) at birth ($\ln(M_{\text{birth}})$). Insets, binned data. **f**, Data from **e** (solid blue dots), supplemented with data from unusually small wild-type haploid daughters (open green circles), generated using essentially the method of ref. 17. For statistical analysis and estimated slopes, see Supplementary Information.

Efficient size control ensures that all cells bud at the same size. As there is variability in cell size at birth, an efficient sizer would ensure that smaller cells spend longer in G1, generating cell-to-cell variability in G1 duration. Measuring individual growth rates and cell sizes allows G1 variability to be decomposed into variability due to size control and a size-independent residual that is attributable to molecular noise. Assuming that G1 duration for an individual cell is the sum of a deterministic function of cell size at birth, $f(M_{\text{birth}})$, and a stochastic variable, η , then: $\alpha T_{\text{G1}} = f(M_{\text{birth}}) + \eta$, where $f(M_{\text{birth}})$ is obtained empirically by binning data. For a measured distribution of sizes at birth, the variance of $f(M_{\text{birth}})$ yields the amount of G1 variability that is produced by size control. Size-independent variability is the average distance between a data point and the deterministic $f(M_{\text{birth}})$.

This decomposition shows that size-independent (presumably molecular) noise is the leading source of variability in the duration of G1 in both mothers and daughters (Table 1). However, size control accounts for about 30–40% of overall G1 variability in daughters. Consistent with our previous timing analysis, size-independent noise decreases by about a factor of $\sqrt{2}$ for each ploidy doubling (Table 1; Supplementary Figs 7, 9, 10); thus, ploidy-dependent noise reduction is robust to statistical removal of all detectable size and growth rate effects, indicating that molecular noise is likely to explain size-independent variability. The dosage of the genes for G1 cyclins also decreases size-independent variability; therefore, some of this variability might be attributable to stochastic variation in the expression of G1 cyclins themselves, although other targets cannot be excluded.

We can divide G1 into two steps on the basis of nuclear residence of Whi5^{24,28}, which enters the nucleus late in mitosis and exits during G1²⁸ (Fig. 3a, b). Whi5 represses the transcription of genes that are activated by the SBF/MBF transcription factors^{28–30}, including *CLN1* and *CLN2*. Cln3 initiates the inactivation and nuclear exit of Whi5^{28–30}. We call the period between cytokinesis and Whi5 exit T_1 , and the period between Whi5 exit and budding T_2 . In mother cells, T_1 is only ~ 1 min (Table 2), so the ~ 15 min length of G1 in mothers is mainly determined by T_2 . By contrast, in daughter cells T_1 and T_2 are about 20 and 17 min, respectively. As T_1 and T_2 are uncorrelated (Fig. 3c), Whi5 exit marks the boundary between two independent steps. For daughters, plotting αT_1 , the time from birth to Whi5 nuclear exit scaled with the growth rate α , against the logarithm of the size at birth, yields a nearly identical relationship to that for overall G1 duration αT_{G1} (Figs 2e and 3d, e), shifted down owing to growth during T_2 . The indicated two-slope model fits these data

Table 1 | Decomposition of G1 variability into a deterministic size control term and a residual attributable to molecular noise

	G1 noise, birth to budding (coefficient of variation of αT_{G1})	Noise due to size control (percentage of the total variance of αT_{G1})	Size and growth rate-independent noise (percentage of the total variance of αT_{G1})
Daughters			
Wild-type haploids	0.55 ± 0.06	0.31 ± 0.04 (32)	0.45 ± 0.04 (68)
Wild-type diploids	0.42 ± 0.04	0.28 ± 0.03 (45)	0.31 ± 0.03 (55)
Wild-type tetraploids	0.24 ± 0.02	0.15 ± 0.01 (39)	0.19 ± 0.02 (61)
Haploid 6 \times CLN2	0.48 ± 0.04	0.30 ± 0.03 (39)	0.37 ± 0.03 (61)
Haploid 6 \times CLN3	0.44 ± 0.04	0.25 ± 0.02 (32)	0.36 ± 0.03 (68)
Haploid 6 \times CLN3 6 \times CLN2	0.37 ± 0.03	0.18 ± 0.02 (24)	0.32 ± 0.03 (76)
Mothers			
Wild-type haploids	0.50 ± 0.05	0.20 ± 0.02 (16)	0.46 ± 0.04 (84)
Wild-type diploids	0.39 ± 0.04	0.13 ± 0.01 (11)	0.37 ± 0.04 (89)
Wild-type tetraploids	0.26 ± 0.02	0.09 ± 0.01 (12)	0.24 ± 0.02 (88)
Haploid 6 \times CLN2	0.33 ± 0.03	0.13 ± 0.01 (16)	0.30 ± 0.03 (84)
Haploid 6 \times CLN3	0.48 ± 0.05	0.16 ± 0.02 (11)	0.45 ± 0.04 (89)
Haploid 6 \times CLN3 6 \times CLN2	0.34 ± 0.03	0.17 ± 0.02 (25)	0.29 ± 0.02 (75)

G1 noise (first column): coefficient of variation of αT_{G1} (α , growth rate; T_{G1} , G1 duration). G1 noise is decomposed into size-dependent and size-independent components (second and third columns); in parentheses, the percentage of the variance of αT_{G1} accounted for in each column. Noise in αT_{G1} is the square root of the sum of the squares of the two independent noise contributions.

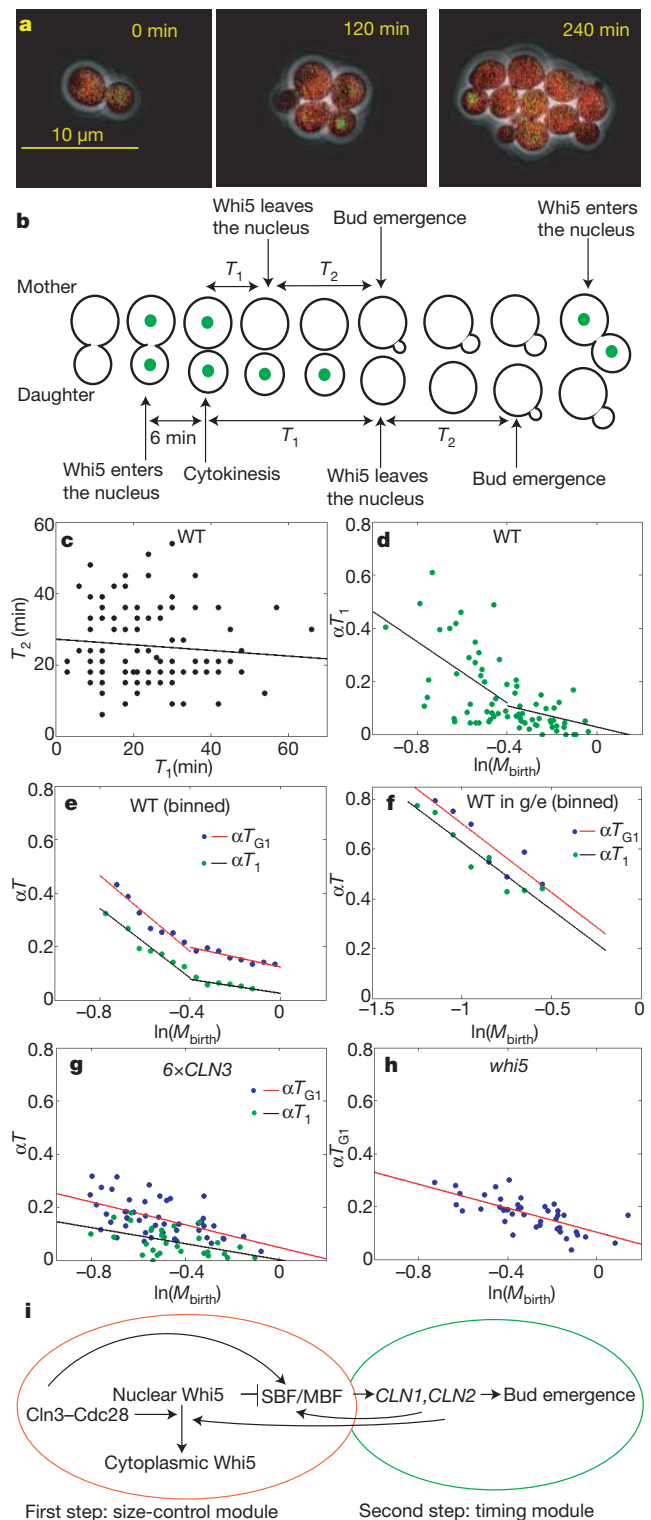


Figure 3 | The correlation between cell size and the regulation of Whi5 nuclear residence supports decomposition of Start into a size-control module and an independent timing module. **a**, Composite phase contrast, Whi5–GFP and *ACT1pr*–DsRed images for haploid cells. Whi5–GFP is mostly observed in newborn daughter cells. **b**, Diagram of the measured intervals; T_{G1} (Fig. 1) is approximately $T_1 + T_2$. **c–h** show only data for daughters. **c**, T_1 and T_2 are uncorrelated (correlation coefficient -0.1) (data from ref. 24; WT, wild-type). **d**, αT_1 against $\ln(M_{\text{birth}})$. **e**, Binned data from Fig. 3d (green points, black line) and Fig. 2e (blue points, red line). **f**, Binned data for αT_1 (green points, black line) and αT_{G1} (blue points, red line) against $\ln(M_{\text{birth}})$ for cells grown in glycerol/ethanol (g/e; see also Supplementary Fig. 3). **g**, αT_1 and αT_{G1} against $\ln(M_{\text{birth}})$ for 6 \times CLN3 cells. **h**, αT_{G1} against $\ln(M_{\text{birth}})$ for *whi5* cells. **i**, Model decomposing Start into a size-control module and an independent timing module unaffected by cell size. For statistical analysis, see Supplementary Information.

Table 2 | Average durations of T_1 and T_2 in different strains

	Wild type	<i>cln3</i>	$6\times CLN3$	$6\times CLN2$
T_1 in daughters	20 ± 1 (157)	30 ± 4 (47)	10 ± 1 (53)	16 ± 1 (80)
T_1 in mothers	0.9 ± 0.3 (170)	13 ± 1 (55)	0.7 ± 0.5 (56)	1.1 ± 0.4 (90)
T_2 in daughters	17 ± 2 (157)	14 ± 2 (47)	16 ± 1 (53)	11 ± 1 (80)
T_2 in mothers	14.7 ± 0.6 (170)	13 ± 1 (55)	13.6 ± 0.8 (56)	12.0 ± 0.6 (90)

T_1 , period from cytokinesis to Whi5 nuclear exit; T_2 , period from Whi5 exit to bud emergence (see Fig. 3b). Mean \pm s.e.m. in min (number of observations in parentheses).

significantly better than a one-slope model, and the deduced slopes for the Whi5 data and for the total G1 data are similar (see Supplementary Information). Thus, G1 size control is restricted to T_1 , the period of Whi5 nuclear residence. T_2 , the part of G1 after Whi5 exit, is independent of cell size, and similar in mothers and daughters²⁴ (Table 2).

These observations are robust to changes in nutrient conditions: growth of cells in glycerol/ethanol instead of glucose, resulting in slow growth and generation of very small newborn daughter cells, gave quantitatively similar results (Fig. 3f; see Supplementary Information). As for the small daughters produced by transient *MET3-CLN2* expression (Fig. 2f), the combined data sets for glycerol/ethanol and glucose strongly supported a two-slope model for total daughter G1 and Whi5 nuclear residence times (see Supplementary Information). Even in glycerol/ethanol, mother cells showed little or no size control (Supplementary Fig. 6).

Increasing *CLN3* gene dosage decreases T_1 , whereas increasing *CLN2* gene dosage modestly decreases T_2 (Table 2, see Supplementary Information). Deletion of *CLN3* significantly increased T_1 in both mothers and daughters, so the shortness of T_1 in wild-type mothers is *Cln3*-dependent (Table 2). Efficient size control, indicated by the steep slope for small wild-type daughters, is essentially eliminated by an increase in *CLN3* gene dosage or deletion of *WHI5* (Fig. 3e, g, h).

Changing the copy numbers of *CLN3* as compared with *CLN2* had differential effects on G1 variability in mothers and daughters (Fig. 1). The two-step model explains this, as an increase in *CLN3* copy number should only affect the first step, which is slow in daughters but very rapid in mothers. As, in mothers, G1 is temporally dominated by the second step, mother cell G1 variability is more sensitive to changes in *CLN2* copy number (Fig. 1). Consistent with this idea and with independence of the two steps, combining $6\times CLN3$ together with $6\times CLN2$ in one haploid genome resulted in low G1 variability in both mothers and daughters (Table 1, Supplementary Fig. 8).

Therefore, we decompose G1 into two independent steps separated by Whi5 nuclear exit: a size-sensing module and a size-independent timing module. The first step depends on both *Cln3* and cell size, and the second step depends on *Cln2*, but not on cell size or *Cln3* (Fig. 3i). Temporal variability in the first step is due to the natural variability in cell size at birth coupled with size control, as well as molecular noise, possibly due to variability in *CLN3* expression. The duration of the second step is cell-size independent; its variability is affected by the expression of the G1 cyclin *CLN2*, one of the primary final effectors of Start^{16–19}. This analysis demonstrates that molecular noise has a role in generating variability in a cellular transition, and provides a precise quantitative framework for the analysis of size control.

METHODS SUMMARY

Strain and plasmid constructions. Standard methods were used throughout. All strains are W303-congenic. All integrated constructs were characterized by Southern blot analysis.

Time-lapse microscopy. Cells were prepared for time-lapse microscopy as described²⁴. We observed growth of microcolonies with fluorescence time-lapse microscopy at 30 °C using a Leica DMIRE2 inverted microscope with a Ludl motorized XY stage. Images were acquired every 3 min for cells grown in glucose and every 6 min for cells grown in glycerol/ethanol with a Hamamatsu Orca-ER

camera. We used custom Visual Basic software integrated with ImagePro Plus to automate image acquisition and microscope control.

Image analysis. Automated image segmentation and fluorescence quantification of yeast grown under time-lapse conditions, and semiautomated assignment of microcolony pedigrees, were performed as described²⁴. Budding was scored visually by the appearance of Myo1-GFP at the incipient bud neck, and division by its disappearance, generally with single-frame accuracy. The nuclear residence of Whi5-GFP was scored by visual inspection of composite phase contrast-fluorescent movies. Cell size was measured as the total cell fluorescence from DsRed protein, expressed from the constitutively active *ACT1* promoter. Background was measured as the average fluorescence of unlabelled cells for each movie and subtracted from the measured pixel intensities. To combine data from different experiments, we normalized our measurements of red fluorescence to the average red fluorescence at budding computed for each colony separately. Similar results were obtained by normalizing the size of cells by using the average intensity computed for each colony separately.

Data analysis. Time-lapse microscopy data were analysed with custom software written in MATLAB. Binning was performed by marking off nonoverlapping intervals on the horizontal axis, averaging the data in each interval and plotting in the centre of the interval.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 March; accepted 6 July 2007.

1. Samoilov, M. S., Price, G. & Arkin, A. P. From fluctuations to phenotypes: the physiology of noise. *Sci. STKE* **366**, re17 (2006).
2. Hartwell, L. H., Culotti, J., Pringle, J. R. & Reid, B. J. Genetic control of the cell division cycle in yeast. *Science* **183**, 46–51 (1974).
3. Hartwell, L. H. & Unger, M. W. Unequal division in *Saccharomyces cerevisiae* and its implications for the control of cell division. *J. Cell Biol.* **75**, 422–435 (1977).
4. Johnston, G. C., Pringle, J. R. & Hartwell, L. H. Coordination of growth with cell division in the yeast *Saccharomyces cerevisiae*. *Exp. Cell Res.* **105**, 79–98 (1977).
5. Moore, S. A. Kinetic evidence for a critical rate of protein synthesis in the *Saccharomyces cerevisiae* yeast cell cycle. *J. Biol. Chem.* **263**, 9674–9681 (1988).
6. Lord, P. G. & Wheals, A. E. Variability in individual cell cycles of *Saccharomyces cerevisiae*. *J. Cell Sci.* **50**, 361–376 (1981).
7. Wheals, A. E. Size control-models of *Saccharomyces cerevisiae* cell proliferation. *Mol. Cell. Biol.* **2**, 361–368 (1982).
8. Lord, P. G. & Wheals, A. E. Rate of cell cycle initiation of yeast cells when cell size is not a rate-determining factor. *J. Cell Sci.* **59**, 183–201 (1983).
9. Nurse, P. Cell cycle control—both deterministic and probabilistic. *Nature* **286**, 9–10 (1980).
10. Spudis, J. L. & Koshland, D. E. Non-genetic individuality: chance in the single cell. *Nature* **262**, 467–471 (1976).
11. Jorgensen, P. & Tyers, M. How cells coordinate growth and division. *Curr. Biol.* **14**, R1014–R1027 (2004).
12. Schroedinger, E. *What is Life?* (Cambridge Univ. Press, Cambridge, 1944).
13. Bi, E. *et al.* Involvement of an actomyosin contractile ring in *Saccharomyces cerevisiae* cytokinesis. *J. Cell Biol.* **142**, 1301–1312 (1998).
14. Cross, F. R. *DAF1*, a mutant gene affecting size control, pheromone arrest, and cell cycle kinetics of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8**, 4675–4684 (1988).
15. Nash, R., Tokiwa, G., Anand, S., Erickson, K. & Futcher, A. B. The WHI1+ gene of *Saccharomyces cerevisiae* tethers cell division to cell size and is a cyclin homolog. *EMBO J.* **7**, 4335–4346 (1988).
16. Tyers, M., Tokiwa, G. & Futcher, B. Comparison of the *Saccharomyces cerevisiae* G1 cyclins: *Cln3* may be an upstream activator of *Cln1*, *Cln2* and other cyclins. *EMBO J.* **12**, 1955–1968 (1993).
17. Dirick, L., Bohm, T. & Nasmyth, K. Roles and regulation of *Cln-Cdc28* kinases at the start of the cell cycle of *Saccharomyces cerevisiae*. *EMBO J.* **14**, 4803–4813 (1995).
18. Stuart, D. & Wittenberg, C. *CLN3*, not positive feedback, determines the timing of *CLN2* transcription in cycling cells. *Genes Dev.* **9**, 2780–2794 (1995).
19. Cross, F. R. Starting the cell cycle: what's the point? *Curr. Opin. Cell Biol.* **7**, 790–797 (1995).
20. McNerny, C. J., Partridge, J. F., Mikesell, G. E., Creemer, D. P. & Breeden, L. L. A novel *Mcm1*-dependent element in the *SWI4*, *CLN3*, *CDC6*, and *CDC47* promoters activates M/G1-specific transcription. *Genes Dev.* **11**, 1277–1288 (1997).
21. Koch, C., Schleiffer, A., Ammerer, G. & Nasmyth, K. Switching transcription on and off during the yeast cell cycle: *Cln/Cdc28* kinases activate bound transcription factor *SBF* (*Swi4/Swi6*) at start, whereas *Cln/Cdc28* kinases displace it from the promoter in G2. *Genes Dev.* **10**, 129–141 (1996).
22. Wijnen, H., Landman, A. & Futcher, B. The G(1) cyclin *Cln3* promotes cell cycle entry via the transcription factor *Swi6*. *Mol. Cell. Biol.* **22**, 4402–4418 (2002).
23. Holstege, F. C. P. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).

24. Bean, J. M., Siggia, E. D. & Cross, F. R. Coherence and timing of cell cycle start examined at single-cell resolution. *Mol. Cell* **21**, 3–14 (2006).
25. Elliott, S. G. & McLaughlin, C. S. Rate of macromolecular synthesis through the cell cycle of the yeast *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **75**, 4384–4388 (1978).
26. Donnan, L. & John, P. C. Cell cycle control by timer and sizer in *Chlamydomonas*. *Nature* **304**, 630–633 (1983).
27. Sveiczer, A., Novak, B. & Mitchinson, J. M. The size control of fission yeast revisited. *J. Cell Sci.* **109**, 2947–2957 (1996).
28. Costanzo, M. *et al.* CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. *Cell* **117**, 899–913 (2004).
29. de Bruin, R. A. M., McDonald, W. H., Kalashnikova, T. I., Yates, J. III & Wittenberg, C. Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. *Cell* **117**, 887–898 (2004).
30. Wittenberg, C. & Reed, S. I. Cell cycle-dependent transcription in yeast: promoters, transcription factors, and transcriptomes. *Oncogene* **24**, 2746–2755 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Haber, E. Bi and the NCRR Yeast Resource Center, University of Washington, for plasmids and strains; P. Nurse and T. Ryan for discussions; and N. Buchler, B. Drapkin, J. Robbins, J. Roberts and J. Widom for comments on the manuscript. This work was supported by the National Institute of Health (J.M.S., E.D.S., F.R.C.), the Howard Hughes Medical Institute (J.M.B.), and the National Science Foundation (E.D.S.).

Author Contributions Experimental work by S.D. and J.M.S.; project planning, data analysis and manuscript preparation by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.R.C. (fcross@mail.rockefeller.edu).

METHODS

Strain and plasmid constructions. Standard methods were used throughout. All strains are of the W303 background. The plasmid pSD03 (pRS403-*CLN2*) was obtained by cloning the *Sma*I-*Sfo*I fragment containing *CLN2* genomic DNA obtained from the *Yep24-CLN2* 2 μ plasmid (J. McKinney, unpublished data) at the *Sma*I site in pRS403. The *CLN2* genomic fragment started 1.4 kb upstream of the *CLN2* open reading frame and ended about 8 kb downstream. A homologous recombination pop-out of the Ty1 (retrotransposon) downstream of *CLN2* was found in the original *Yep24-CLN2* 2 μ plasmid. pSD03 was integrated at the *HIS3* locus by *Bgl*II digestion. Strain SD27-1-1A was confirmed by Southern blot analysis to have five duplicative integrations of pSD03. The plasmid pJB06T (pRS404-*ACT1pr-DSRED*) was obtained as follows. The *ACT1pr* (~500 bp upstream of the *ACT1* open reading frame) was inserted into pTY24 (obtained from NCRR Yeast Resource Center, University of Washington) just upstream of *DSRED* coding sequence. The *Bam*HI-*Bgl*II fragment containing *ACT1pr-DSRED* was then inserted at the *Bam*HI site in pRS404. Plasmid pSD02 (pRS406-*ACT1pr-DSRED*) was obtained by cloning the *Bgl*II fragment containing *ACT1pr-DSRED* obtained from pJB06T into the *Bgl*II fragment containing *URA3* obtained from pRS406. pSD02 was integrated at the *URA3* locus by *Nco*I digestion. Strain SD20-1A was confirmed by Southern blot analysis to have two duplicative integrations of pSD02. All the other strains that carry pSD02 were obtained by crosses with SD20-1A or with strains derived from it so they also have two duplicative integrations of pSD02. We observed that strains transformed with one copy of pJB06T or two copies of pSD02 behave identically with the only difference that the average intensity of the second reporter is about twice as large as the average intensity of the first one, as expected by difference in copy number. A *MYO1-GFP* strain¹³ was backcrossed at least five times to W303 to obtain the strains used in this paper.

Time-lapse microscopy. Cells were prepared for microscopy and time-lapse microscopy as described²⁴. We observed growth of microcolonies with fluorescence time-lapse microscopy at 30 °C using a Leica DMIRE2 inverted microscope with a Ludl motorized XY stage. Images were acquired every 3 min for cells grown in glucose and every 6 min for cells grown in glycerol/ethanol with a

Hamamatsu Orca-ER camera. We used custom Visual Basic software integrated with ImagePro Plus to automate image acquisition and microscope control.

Image analysis. Automated image segmentation and fluorescence quantification of yeast grown under time-lapse conditions, and semiautomated assignment of microcolony pedigrees, were performed as described²⁴. Budding and division were scored by visual inspection for the appearance and disappearance of the Myo1-GFP signal, respectively. Detection of the appearance of the Myo1-GFP signal was facilitated by thresholding the images and plotting the pixels above threshold on a linear grey scale. The nuclear residence of Whi5-GFP was scored by visual inspection of composite phase contrast-fluorescent movies and confirmed by the method described for Myo1-GFP detection. Scoring of cell cycle events by monitoring the localization of Myo1-GFP and Whi5-GFP is much faster than detection of newly synthesized, cell-cycle-regulated fluorescent proteins and allows a much higher temporal resolution. Cell size was measured as the total cell fluorescence from DsRed protein, expressed from the constitutively active *ACT1* promoter. Background was measured as the average fluorescence of unlabelled cells for each movie and subtracted from the measured pixel intensities. The measurement of cell size at given time points displayed an appreciable variability, probably due to experimental noise in the imaging and errors in the segmentation of cell bodies. The effect of this noise on size measurements was efficiently averaged out by extracting cell size at a given time point from the least-square fit to exponential growth. The average fluorescent intensity of identical strains imaged on different days and with different fluorescent lamps displayed variability much larger than natural variability among cells in the same colony. To combine data from different experiments we normalized our measurements of cell size to the average size at budding computed for each colony separately. Similar results were obtained by normalizing the size of cells by using the average intensity computed for each colony separately.

Data analysis. We analysed time-lapse microscopy timing and fluorescent data with custom software written in MATLAB. Binning was performed as follows: multiple points were placed in bins and averaged. The averages and the values of the centre of the bin were then plotted. Similar results to that obtained by binning the data were obtained by smoothing the data with a moving average.

LETTERS

Antidepressant binding site in a bacterial homologue of neurotransmitter transporters

Satinder K. Singh¹, Atsuko Yamashita^{3†} & Eric Gouaux^{1,2}

Sodium-coupled transporters are ubiquitous pumps that harness pre-existing sodium gradients to catalyse the thermodynamically unfavourable uptake of essential nutrients, neurotransmitters and inorganic ions across the lipid bilayer¹. Dysfunction of these integral membrane proteins has been implicated in glucose/galactose malabsorption², congenital hypothyroidism³, Bartter's syndrome⁴, epilepsy⁵, depression⁶, autism⁷ and obsessive-compulsive disorder⁸. Sodium-coupled transporters are blocked by a number of therapeutically important compounds, including diuretics⁹, anticonvulsants¹⁰ and antidepressants¹¹, many of which have also become indispensable tools in biochemical experiments designed to probe antagonist binding sites and to elucidate transport mechanisms. Steady-state kinetic data have revealed that both competitive^{12,13} and noncompetitive^{14,15} modes of inhibition exist. Antagonist dissociation experiments on the serotonin transporter (SERT) have also unveiled the existence of a low-affinity allosteric site that slows the dissociation of inhibitors from a separate high-affinity site¹⁶. Despite these strides, atomic-level insights into inhibitor action have remained elusive. Here we screen a panel of molecules for their ability to inhibit LeuT, a prokaryotic homologue of mammalian neurotransmitter sodium symporters, and show that the tricyclic antidepressant (TCA) clomipramine noncompetitively inhibits substrate uptake. Cocystal structures show that clomipramine, along with two other TCAs, binds in an extracellular-facing vestibule about 11 Å above the substrate and two sodium ions, apparently stabilizing the extracellular gate in a closed conformation. Off-rate assays establish that clomipramine reduces the rate at which leucine dissociates from LeuT and reinforce our contention that this TCA inhibits LeuT by slowing substrate release. Our results represent a molecular view into noncompetitive inhibition of a sodium-coupled transporter and define principles for the rational design of new inhibitors.

LeuT is a stable, sodium-coupled leucine transporter from the eubacterium *Aquifex aeolicus* and is the only member of the neurotransmitter sodium symporter (NSS or SLC6) family of secondary transporters that has so far been amenable to structural analysis¹⁷. Eukaryotic NSS counterparts include those that pump neurotransmitters such as serotonin, norepinephrine, dopamine, glycine and γ -aminobutyric acid (GABA) from the synapse to neuronal and glial cytoplasm, shaping the magnitude and duration of synaptic signalling¹⁸. The crystal structure of LeuT provided the first molecular glimpse into an NSS member¹⁷, but it did not yield any clues as to the atomic basis of inhibition. Motivated by this dearth of knowledge, together with the key role of NSS proteins and their antagonists in synaptic transmission, we sought to find inhibitors of LeuT and to characterize their mechanism of action by using flux, binding and crystallographic methods.

We began by screening a wide variety of NSS inhibitors, including those that target the transporters of glycine (GlyT), GABA (GAT), dopamine (DAT), norepinephrine (NET), and serotonin (SERT), for their ability to inhibit the uptake of L-[³H]Leu by purified LeuT reconstituted into lipid vesicles. The tricyclic antidepressant (TCA) clomipramine was the most potent inhibitor of uptake (Fig. 1a), whereas the structurally related compounds, imipramine and desipramine, were less effective (Fig. 1a, Supplementary Fig. 1a). In dose-response experiments (Fig. 1b), clomipramine was a modestly potent inhibitor of leucine transport, with a half-maximal inhibitory concentration (IC₅₀) approximately eightfold lower than that of imipramine (250 versus 2,090 μ M; Supplementary Table 1). To test whether the TCAs could displace bound leucine, we conducted radioligand binding experiments using detergent-solubilized LeuT, [³H]Leu, and potential cold competitors. Unlike amino acids such as leucine, alanine and tryptophan, none of the TCAs (at 1 mM) significantly displaced [³H]Leu from LeuT (Supplementary Fig. 1b), tentatively excluding competitive inhibition as the mechanism by which the TCAs inhibit LeuT.

To define the mode of inhibition by clomipramine more precisely, we next performed a series of steady-state kinetic experiments. With [³H]Leu as the substrate, clomipramine unambiguously reduced the maximum velocity of uptake at infinite concentration of substrate (V_{\max}), although there was no clear correlation with the Michaelis constant (K_m ; Supplementary Fig. 1c). This might be because at 20 °C LeuT exhibits a high affinity for leucine (dissociation constant (K_d) \approx 20 nM) and turns it over slowly, with a catalytic rate constant (k_{cat}) of only 1.2 h⁻¹ (S.K.S., C. Piscitelli, A.Y. & E.G., manuscript in preparation), thus obscuring the distinction between simple binding and transport in a standard flux assay (Supplementary Fig. 1d). However, L-alanine is also a substrate of LeuT (Supplementary Fig. 1d), it binds with a K_d at least 15-fold greater than that of leucine, and it has k_{cat} and K_m values \sim 5-fold greater and \sim 3–4-fold greater, respectively, than the corresponding values for leucine (S.K.S., C.P., A.Y. & E.G., manuscript in preparation). Importantly, clomipramine is a more potent inhibitor of [³H]Ala transport with an IC₅₀ of \sim 5 μ M (Supplementary Fig. 1e, Supplementary Table 1). With [³H]Ala as the substrate and clomipramine as the inhibitor, the Michaelis–Menten plot (Supplementary Fig. 1f) showed a decrease in V_{\max} with no change in K_m , and the Eadie–Hofstee graph (Fig. 1c) had parallel lines with increasing clomipramine concentrations, both unequivocal indicators of noncompetitive inhibition (Supplementary Table 1). Crystallographic analysis confirmed that the LeuT–alanine–sodium complex adopts the same occluded state as the leucine complex (PDB code 2A65)¹⁷, except for minute alterations in the substrate binding pocket (S.K.S., C.P., A.Y. & E.G., manuscript in preparation), validating the use of [³H]Ala in the kinetic experiments.

¹The Vollum Institute and ²Howard Hughes Medical Institute, Oregon Health and Science University, 3181 S.W. Sam Jackson Road, Portland, Oregon 97239, USA. ³Department of Biochemistry and Molecular Biophysics, Columbia University, 650 West 168th Street, New York, New York 10032, USA. [†]Present address: RIKEN SPring-8 Center, 1-1-1, Kouto, Sayo, Hyogo 679-5148, Japan.

Where does clomipramine bind and what is the molecular basis for its inhibition of substrate transport? To address these questions, we solved cocrystal structures of the leucine- and sodium-bound state of LeuT in complex with clomipramine, desipramine or imipramine as well as the alanine- and sodium-bound state of LeuT in complex with clomipramine. All four crystal structures refined well ($R_{\text{free}} = 21.0\text{--}21.9\%$) using data measured to $1.7\text{--}1.9\text{ \AA}$ resolution (Supplementary Table 2). Simulated-annealing $F_o - F_c$ omit maps revealed the saddle-shaped electron density expected for the dibenzazepine ring, and for clomipramine there was a clear protrusion at the 3-position

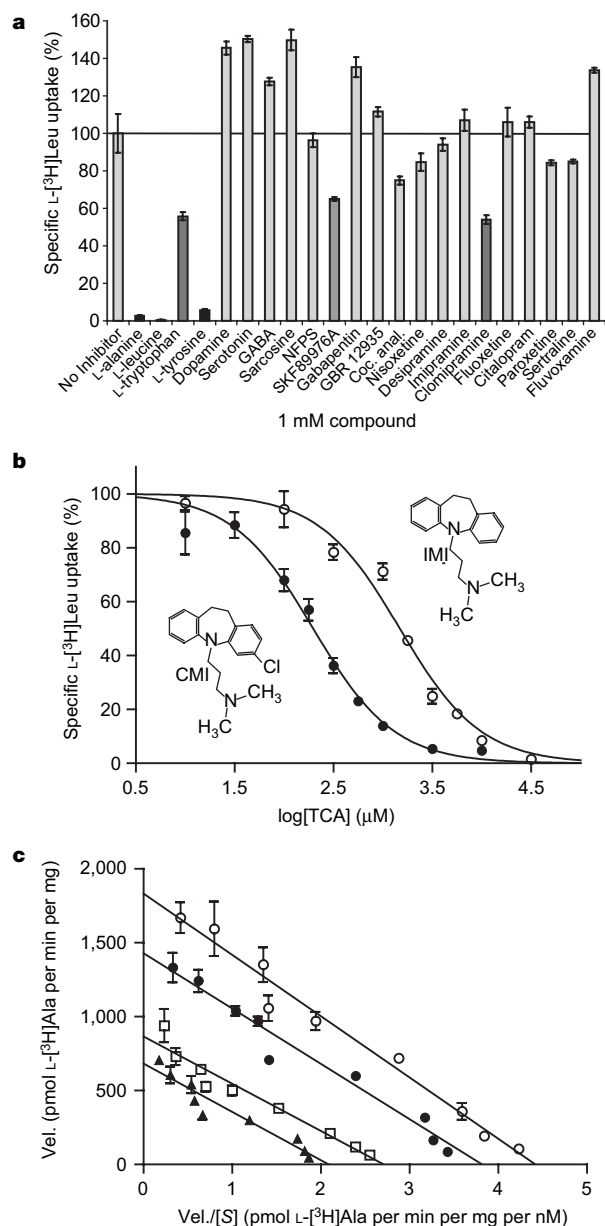


Figure 1 | Functional analysis of LeuT and effect of inhibitors and TCAs. **a**, Inhibition of L-[³H]leucine transport by a panel of eukaryotic NSS substrates and inhibitors, all at 1 mM. NFPS, *N*[3-(4'-fluorophenyl)-3-(4'-phenylphenoxy) propyl] sarcosine; Coc. anal., cocaine analogue 3- α -(4-chlorophenyl)phenylmethoxy]tropane. **b**, Dose dependence of inhibition by clomipramine (CMI; filled circles) and imipramine (IMI; open circles) of [³H]Leu transport. **c**, Steady-state kinetics of inhibition by clomipramine of L-[³H]alanine transport (Eadie-Hofstee plot). Open circles, filled circles, open squares and filled triangles represent data for 0, 1, 20 and 50 μM clomipramine, respectively. [S] is the substrate concentration and vel. is the velocity of substrate uptake. Error bars represent s.e.m. of triplicate measurements.

pinpointing the location of the chlorine atom (Fig. 2a), a feature absent in the density associated with both imipramine (Fig. 2b) and desipramine (data not shown). Clomipramine (Fig. 2c), along with the other TCAs (Supplementary Fig. 2), binds to LeuT in the extracellular-facing vestibule that probably forms part of the

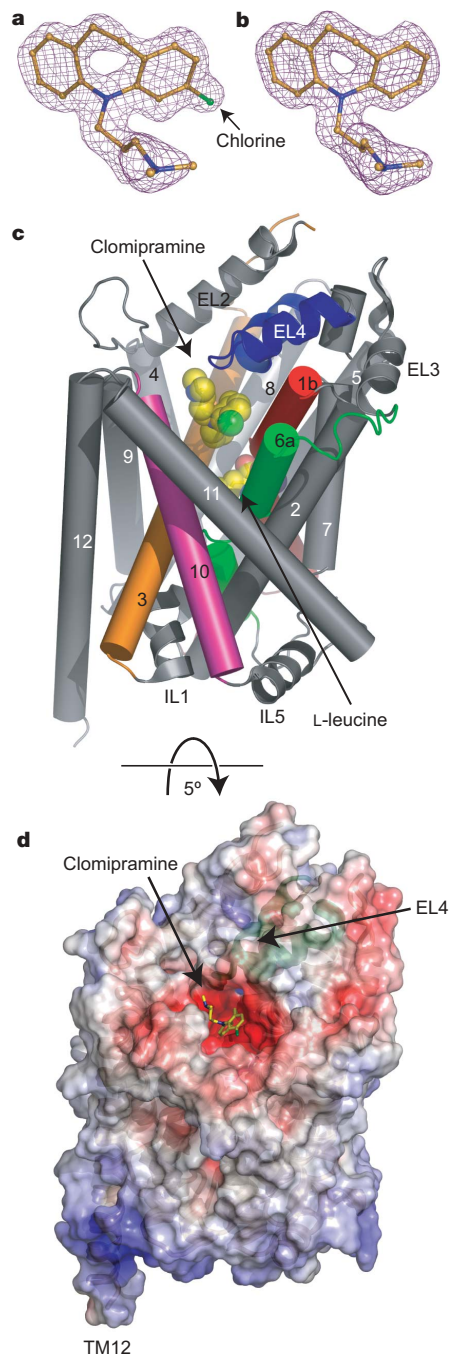


Figure 2 | TCAs bind in the putative permeation pathway of LeuT. $F_o - F_c$ simulated annealing omit maps, both contoured at 3.0σ , of clomipramine (**a**) and imipramine (**b**), in which the TCA was omitted from the simulated annealing run and subsequent phase calculation. The chlorine atom of clomipramine is green. **c**, LeuT tilted $\sim 15^\circ$ from the membrane plane to illustrate the binding sites of clomipramine and leucine, both depicted in space-filling representation. Helices whose residues interact with clomipramine are coloured. **d**, Electrostatic properties of the LeuT extracellular-facing vestibule, with clomipramine (yellow) cradled in the negatively charged crevice. Electrostatic potential isocontours were set at $+7\text{ kT e}^{-1}$ (blue) and -7 kT e^{-1} (red). EL4 is tinted green. To make clomipramine more visible, LeuT has been tilted an additional $\sim 5^\circ$ towards the reader from the view in **c** ($\sim 20^\circ$ from the membrane plane).

permeation pathway, lying between the extracellular solution and the occluded substrate- and ion-binding pocket located halfway across the membrane bilayer. The TCAs are cradled by the carboxy-terminal half of transmembrane helix 1 (TM1), the amino-terminal regions of TM6 and TM10, the approximate midpoint of TM3, and the sharp turn of extracellular loop 4 (EL4), about 11 Å above the substrate and two sodium ions. Although most of the residues lining the TCA-binding pocket are nonpolar (Fig. 3a), essentially complementing the hydrophobic character of the TCAs, an electrostatic calculation shows that the binding site has a pronounced negative potential that probably enhances the binding of positively charged inhibitors (Fig. 2d). Clomipramine is further stabilized by a salt bridge between its protonated¹⁹ N2 atom and the side-chain carboxylate (Oδ1) of Asp 401 (2.9 Å) as well as a polar interaction between its chlorine atom and the side-chain amide (Ne2) of Gln 34 (3.1 Å; Fig. 3a). Neither imipramine nor desipramine possesses a chlorine atom and thus they cannot participate in a similar interaction with Gln 34, which might partly explain why they are less potent inhibitors. There is a second, weak TCA binding site on the lipid-exposed face of the 'V' formed between TM4 and TM5, near the intracellular surface (Supplementary Fig. 3), but because this site has a low occupancy and is situated far from the substrate-binding site and probable transport pathway, we believe it is not functionally relevant.

The conformation of the LeuT–TCA complexes is similar to that determined in the original complex with two crucial exceptions.

The first is localized to the guanidium group of Arg 30, an amino acid which, along with Asp 404, forms a highly conserved charged pair that comprises part of the proposed extracellular gate. In the LeuT–leucine–sodium complex, this arginine and aspartate interact indirectly via two water molecules. In the TCA complexes, the guanidium group forms a direct salt bridge with Asp 404, displacing the intervening water molecules (Fig. 3a, b, d). The TCAs now completely shield the salt bridge between Asp 404 and Arg 30 from aqueous solution, thereby lowering the local dielectric constant and presumably increasing the strength of the ion pair. The new position of the Arg 30 guanidium group is further buttressed by a cation- π interaction with Phe 253 (Fig. 3e), a residue that directly contacts the bound substrate. Significantly, the flipped position of the guanidium group does not prevent Arg 30 from engaging in a hydrogen-bonding network that ultimately connects it to Na1 and the carboxylate of the substrate, thereby physically linking the TCA- and substrate-binding sites. Intervening atoms include one of the displaced water molecules, Oε1 and Ne2 of Gln 250, Oε1 and Oε2 of Glu 290, and Oγ1 of Thr 254, a series of interactions that might thermodynamically couple the binding of substrate and clomipramine, as supported by our observation that clomipramine exhibits different inhibitory potencies for [³H]Ala and [³H]Leu transport.

The second change is localized to the sharp turn at Ala 319 in EL4 (Fig. 3a, c, d), a region that exhibits conformational mobility during the transport cycle in GAT²⁰, DAT²¹, GlyT1b²² and SERT²³. In the

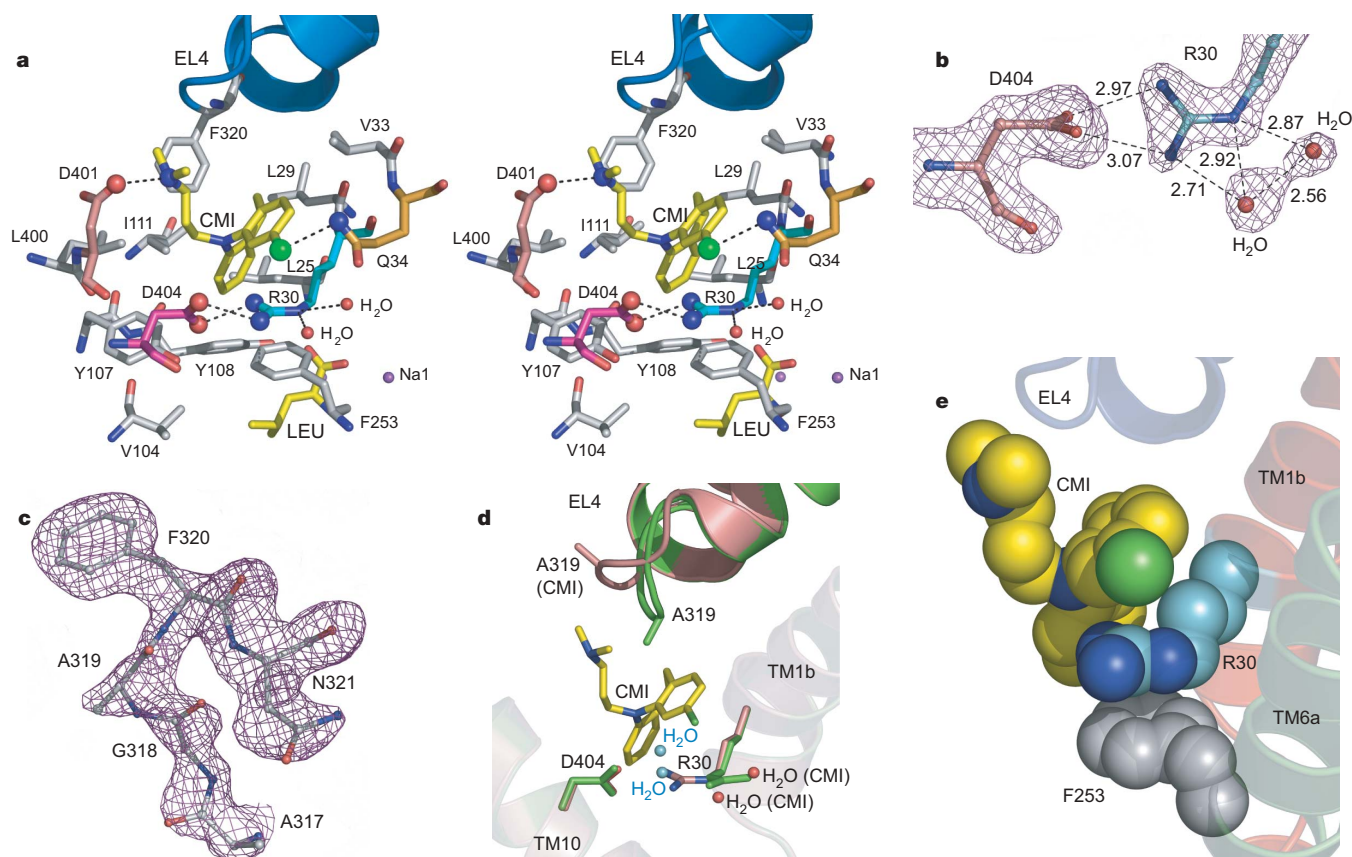


Figure 3 | Clomipramine-binding site. **a**, Stereo view, with clomipramine and leucine carbon atoms depicted in yellow, sodium ions in purple and two water molecules in red. Residues whose interactions with clomipramine are hydrophobic, polar (Q34) or ionic (D401) are coloured grey, orange and pink, respectively. Atoms depicted as spheres interact through either a hydrogen bond or a salt bridge. **b**, $F_o - F_c$ simulated annealing omit map, contoured at 3.0 σ , of D404 and R30, depicting the direct salt bridge between the guanidium of R30 and the carboxylate of D404, with displacement of two water molecules (compare to overlay in **d**). Distances (Å) are shown along dashed lines. **c**, $F_o - F_c$ simulated annealing omit map, contoured at 3.0 σ , of

the tip of EL4 (residues A317–N321), illustrating the movement ‘up’ of A319 in the LeuT–clomipramine crystal structure (compare to overlay in **d**). **d**, Same view as in **a** with residues from the original LeuT structure (PDB ID 2A65) overlaid (in green with two water molecules between D404 and R30 in cyan) onto those from the LeuT–clomipramine crystal structure (in pink with the two displaced waters in red and labelled as H₂O (CMI)). **e**, CPK rendering of clomipramine, R30 and F253 illustrating how the positively charged guanidium group is sandwiched between the aromatic rings of clomipramine and F253.

leucine–sodium complex, the methyl group of Ala 319 points towards the cavity occupied by the TCAs. In the clomipramine complexes, however, Ala 319 is flipped towards the extracellular side of the transporter, to accommodate binding of the inhibitor (Fig. 3c, d). In the imipramine and desipramine complexes, Ala 319 points downwards, as in the leucine–sodium and alanine–sodium complexes, but the residues in the sharp turn (Ala 317–Ala 319) are shifted ‘up’ by as much as 1.4 Å (Supplementary Fig. 4a, b). This shift provides a second mechanism by which movement of EL4 participates in TCA binding in LeuT.

The location of the TCA-binding site in the extracellular-facing cavity of LeuT and on top of residues that have been implicated as part of the extracellular gate suggests a simple mechanism by which these molecules might inhibit substrate transport. We postulate that the occluded state of LeuT, as visualized in the leucine and sodium or the alanine and sodium complexes, is an essential intermediate in the translocation cycle and that the TCAs stabilize this occluded state, thereby trapping the substrate on the transporter and impeding its release to either the intracellular milieu or back to extracellular solution (Fig. 4a). If this hypothesis is correct, then substrate should dissociate from LeuT more slowly in the presence of clomipramine. We used [^3H]Leu for these experiments because [^3H]Ala dissociates too quickly to accurately measure its off-rate (data not shown). A saturating concentration of clomipramine does markedly decrease the off-rate of [^3H]Leu by as much as 700-fold (Fig. 4b, c and Supplementary Table 1). Moreover, fitting the measurements in the absence of clomipramine to a double as opposed to a single exponential yields a statistically superior fit ($P = 0.0003$), perhaps because substrate can

unbind by way of either the extracellular or intracellular pathways. In contrast, the off-rate data in the presence of clomipramine are most adequately fit by a single exponential function ($P = 0.6162$).

How, in molecular detail, does clomipramine slow the release of bound leucine? An explanation can be presented in terms of a variation of the alternating access mechanism²⁴, in which the transporter assumes at least three conformational states: open-to-outside, occluded, and open-to-inside. Access to either side of the membrane bilayer is controlled by extracellular and intracellular gates. In the occluded LeuT structure, in which both gates are closed, access to the substrate-binding site from the extracellular side is blocked by only a few residues, whereas access from the cytoplasmic side is obstructed by ~ 25 Å of ordered protein structure.

Clomipramine and related compounds probably retard the unbinding of substrate through the extracellular pathway by directly stabilizing the salt bridge between Arg 30 and Asp 404, an ionic ‘latch’ that in turn stabilizes key aromatic residues, Tyr 108 and Phe 253, immediately above the bound substrate (Fig. 4d). Clomipramine might also slow the rate of leucine release through the intracellular pathway by allosterically precluding movement of TM1a and TM6b, perhaps together with other intracellular proximal regions such as TM3 and TM8, portions of the transporter that are likely to rearrange on substrate release (Fig. 4d). TM5 and TM10 might also participate in this transition to the cytoplasmic-facing form, as has been suggested from accessibility studies on SERT^{25,26}. We speculate that clomipramine might prevent the extracellular-facing cavity from compacting during interconversion to the intracellular-facing form. Such a condensation in the extracellular-facing cavity has been postulated to occur in

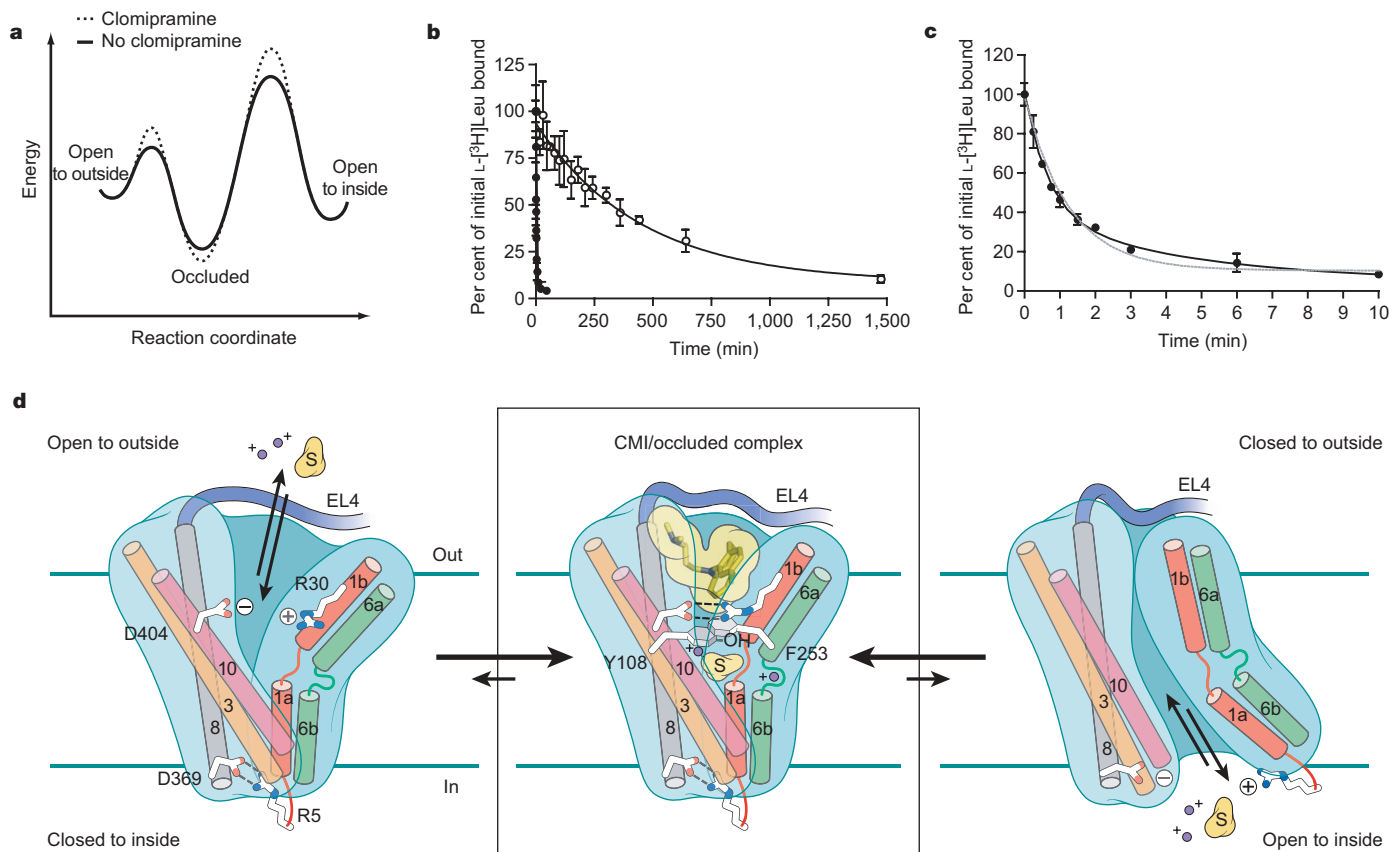


Figure 4 | Clomipramine traps substrate on LeuT. **a**, Hypothetical graph depicting the free energy changes occurring along the transport cycle in the absence (solid lines) and presence (dotted lines) of clomipramine. **b**, Dissociation kinetics of [^3H]Leu from LeuT on a 15-fold dilution into buffer containing 10 mM cold L-leucine and either 0 (closed circles; $k_{\text{off}1} = 85.9 \pm 18.1 \text{ h}^{-1}$ ($62.0 \pm 4.9\%$), $k_{\text{off}2} = 12.1 \pm 4.2 \text{ h}^{-1}$ ($33.5 \pm 4.8\%$)) or 3 mM (open circles; $k_{\text{off}} = 0.12 \pm 0.02 \text{ h}^{-1}$) clomipramine.

c, Magnification of the '0 mM CMI' dissociation reaction, with the single and double exponential fits depicted in grey dotted and black solid lines, respectively. All data to 60 min were included in the fitting, but only the first 10-min segment is displayed. Error bars represent s.e.m. of duplicate measurements. **d**, Schematic illustrating a mechanism for noncompetitive inhibition of LeuT by the TCAs.

SERT^{23,25,27}, where the extracellular inhibitor cocaine is thought to prevent such a compaction from occurring to preclude subsequent opening of the cytoplasmic permeation pathway²⁵.

What relationship do our studies on LeuT have to the mechanism by which the TCAs inhibit NET and SERT, the targets of these drugs in humans? LeuT is clearly related in amino acid sequence to these biogenic amine transporters and, therefore, the three proteins must share a common architecture that probably includes a large extracellular-facing vestibule—a cone-shaped cavity of both hydrophobic and hydrophilic character that can serve as a binding site for molecules as diverse as TCAs and snail peptide toxins²⁸. However, the sequences of SERT and NET still diverge substantially from that of LeuT and, thus, the details of inhibitor binding are likely to differ. Recent data on SERT²⁹ and NET²⁸ indicate that the TCAs probably occupy a primary site that is deeper in the protein core, close to or overlapping with the occluded substrate-binding pocket identified in the LeuT structures. Nevertheless, by demonstrating that molecules bound within the extracellular vestibule can exert noncompetitive inhibition of transport via stabilization of the occluded state, we have pinpointed an additional region that might be exploited for the development of new inhibitors with therapeutic potential.

METHODS

Preparation of LeuT proteoliposomes. LeuT was expressed and purified as described¹⁷, and liposomes were prepared with a 3:1 (w:w) ratio of *Escherichia coli* total lipid extract and egg phosphatidylcholine in buffer I (20 mM HEPES-Tris (pH 7.0), 100 mM potassium gluconate). They were treated with Triton X-100 at a 2.5:1 (w:w) ratio of lipid to detergent before reconstituting with LeuT at protein-to-lipid ratio of 1:110 (w:w) and removing excess detergent with BioBeads. LeuT proteoliposomes were then loaded with buffer I and stored at -80 °C until needed. The efficiency of protein incorporation into lipid vesicles approached 100%, as assessed by a modified amido black assay³⁰. This value was then used in all functional experiments to estimate what quantity of LeuT to use in a specific assay and to convert specific transport activity expressed as pmol [³H]Leu per min to specific activity expressed as pmol [³H]Leu per min per mg protein. Proteolysis experiments established that ~60% of the LeuT is oriented with the N and C termini inside the liposome.

Details of the inhibition screens and time courses as well as the dose response, steady-state inhibition kinetics, and dissociation kinetics experiments are provided in Supplementary Information.

Crystallography. LeuT was expressed and purified as described¹⁷ and then divided into three 30–40-μl aliquots. TCAs were added to a final concentration of 10 mM, incubated on ice for 15 min and then subjected to crystallization trials by hanging drop vapour diffusion. For the LeuT–alanine–clomipramine complex, protein was purified as outlined for the LeuT–leucine–TCA complexes except for the presence of 100 mM alanine in all buffers. Crystals grew within 7–10 days at 18 °C in the presence of 17–22% PEG 550 MME, 100 mM HEPES-NaOH (pH 7–7.5) and 200 mM NaCl. They were cryoprotected by gradually increasing the starting PEG 550 MME concentration to 35%.

Details of data collection and structure determination are provided in Supplementary Information.

Received 30 January; accepted 21 June 2007.

Published online 8 August 2007; corrected 23 August 2007 (details online).

- Hediger, M. A. *et al.* The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. Introduction. *Pflügers Arch.* **447**, 465–468 (2004).
- Wright, E. M., Turk, E. & Martin, M. G. Molecular basis for glucose-galactose malabsorption. *Cell Biochem. Biophys.* **36**, 115–121 (2002).
- Pohlentz, J. & Refetoff, S. Mutations in the sodium/iodide symporter (NIS) gene as a cause for iodide transport defects and congenital hypothyroidism. *Biochimie* **81**, 469–476 (1999).
- Simon, D. B. *et al.* Gitelman's variant of Bartter's syndrome, inherited hypokalaemic alkalosis, is caused by mutations in the thiazide-sensitive Na-Cl cotransporter. *Nature Genet.* **12**, 24–30 (1996).
- Richerson, G. B. & Wu, Y. Role of the GABA transporter in epilepsy. *Adv. Exp. Med. Biol.* **548**, 76–91 (2004).
- Hahn, M. K. & Blakely, R. D. Monoamine transporter gene structure and polymorphisms in relation to psychiatric and other complex disorders. *Pharmacogenomics J.* **2**, 217–235 (2002).
- Sutcliffe, J. S. *et al.* Allelic heterogeneity at the serotonin transporter locus (SLC6A4) confers susceptibility to autism and rigid-compulsive behaviors. *Am. J. Hum. Genet.* **77**, 265–279 (2005).

- Ozaki, N. *et al.* Serotonin transporter missense mutation associated with a complex neuropsychiatric phenotype. *Mol. Psychiatry* **8**, 933–936 (2003).
- Beaumont, K., Vaughn, D. A. & Fanestil, D. D. Thiazide diuretic drug receptors in rat kidney: identification with [³H]metolazone. *Proc. Natl Acad. Sci. USA* **85**, 2311–2314 (1988).
- Krogsgaard-Larsen, P., Frolund, B. & Frydenvang, K. GABA uptake inhibitors. Design, molecular pharmacology and therapeutic aspects. *Curr. Pharm. Des.* **6**, 1193–1209 (2000).
- Barker, E. L. & Blakely, R. D. in *Psychopharmacology—the Fourth Generation of Progress* (eds Bloom, F. E. & Kupfer, D. J.) 321–333 (Raven, New York, 1995).
- Talvenheimo, J., Nelson, P. J. & Rudnick, G. Mechanism of imipramine inhibition of platelet 5-hydroxytryptamine transport. *J. Biol. Chem.* **254**, 4631–4635 (1979).
- Smart, L. Competitive inhibition of sodium-dependent high affinity choline uptake by harmala alkaloids. *Eur. J. Pharmacol.* **75**, 265–269 (1981).
- Vroye, L. *et al.* The Na⁺/I⁻ cotransporter of the thyroid: characterisation of new inhibitors. *Pflügers Arch.* **435**, 259–266 (1998).
- Uryu, K. *et al.* Inhibition by neuromuscular blocking drugs of norepinephrine transporter in cultured bovine adrenal medullary cells. *Anesth. Analg.* **91**, 546–551 (2000).
- Neubauer, H. A., Hansen, C. G. & Wiborg, O. Dissection of an allosteric mechanism on the serotonin transporter: a cross-species study. *Mol. Pharmacol.* **69**, 1242–1250 (2006).
- Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na⁺/Cl⁻-dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).
- Zahniser, N. R. & Doolen, S. Chronic and acute regulation of Na⁺/Cl⁻-dependent neurotransmitter transporters: drugs, substrates, presynaptic receptors, and signaling systems. *Pharmacol. Ther.* **92**, 21–55 (2001).
- Cantu, M. D., Hillebranda, S. & Carrilho, E. Determination of the dissociation constants (pKa) of secondary and tertiary amines in organic media by capillary electrophoresis and their role in the electrophoretic mobility order inversion. *J. Chromatogr. A* **1068**, 99–105 (2005).
- Zomot, E. & Kanner, B. I. The interaction of the γ-aminobutyric acid transporter GAT-1 with the neurotransmitter is selectively impaired by sulfhydryl modification of a conformationally sensitive cysteine residue engineered into extracellular loop IV. *J. Biol. Chem.* **278**, 42950–42958 (2003).
- Norregaard, L., Frederiksen, D., Nielsen, E. O. & Gether, U. Delineation of an endogenous zinc-binding site in the human dopamine transporter. *EMBO J.* **17**, 4266–4273 (1998).
- Ju, P., Aubrey, K. R. & Vandenberg, R. J. Zn²⁺ inhibits glycine transport by glycine transporter subtype 1b. *J. Biol. Chem.* **279**, 22983–22991 (2004).
- Mitchell, S. M., Lee, E., Garcia, M. L. & Stephan, M. M. Structure and function of extracellular loop 4 of the serotonin transporter as revealed by cysteine-scanning mutagenesis. *J. Biol. Chem.* **279**, 24089–24099 (2004).
- Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969–970 (1966).
- Zhang, Y. W. & Rudnick, G. The cytoplasmic substrate permeation pathway of serotonin transporter. *J. Biol. Chem.* **281**, 36213–36220 (2006).
- Keller, P. C. II, Stephan, M., Glomska, H. & Rudnick, G. Cysteine-scanning mutagenesis of the fifth external loop of serotonin transporter. *Biochemistry* **43**, 8510–8516 (2004).
- Henry, L. K., Adkins, E. M., Han, Q. & Blakely, R. D. Serotonin and cocaine-sensitive inactivation of human serotonin transporters by methanethiosulfonates targeted to transmembrane domain I. *J. Biol. Chem.* **278**, 37052–37063 (2003).
- Paczkowski, F. A., Sharpe, I. A., Dutertre, S. & Lewis, R. J. γ-Conotoxin and tricyclic antidepressant interactions at the norepinephrine transporter define a new transporter model. *J. Biol. Chem.* **282**, 17837–17844 (2007).
- Henry, L. K. *et al.* Tyr-95 and Ile-172 in transmembrane segments 1 and 3 of human serotonin transporters interact to establish high affinity recognition of antidepressants. *J. Biol. Chem.* **281**, 2012–2023 (2006).
- Kaplan, R. S. & Pedersen, P. L. Determination of microgram quantities of protein in the presence of milligram levels of lipid with amido black 10B. *Anal. Biochem.* **150**, 97–104 (1985).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. C. LaPorte, R. D. Blakely, and G. Rudnick for insightful comments; M. Post for information on TCA crystallographic numbering; C. Piscitelli for assistance with LeuT purification and electrostatic calculations; L. Vaskalis for Fig. 4a, d; and the staff at beamlines X26C and X29A of the National Synchrotron Light Source and beamline 8.2.2 of the Advanced Light Source for assistance with X-ray data collection. S.K.S. was supported by an individual NIH National Research Service Award. This work was supported by the NIH. E.G. is an investigator with the Howard Hughes Medical Institute.

Author Information Coordinates and structure factors for the alanine–sodium–clomipramine and the leucine–sodium–clomipramine, –imipramine and –desipramine complexes have been deposited in the Protein Data Bank under accession codes 2QE1, 2Q6H, 2Q72 and 2QB4, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.G. (gouauxe@ohsu.edu).

Tissue issues

Millions of tissue samples have been collected and archived, but researchers wanting to explore them at the molecular level have found it tough going. **Nathan Blow** investigates the issues.

According to experts, there are more than a billion tissue samples archived in hospitals and tissue banks around the world, most of them formalin-fixed and paraffin-embedded (FFPE). Today, these samples present both an incredible opportunity and a huge challenge to researchers. FFPE tissue samples have been extensively annotated and well preserved, allowing detailed study of the progression of diseases such as cancer. But due to the method of preservation, obtaining biomolecules from these samples is proving difficult, to say the least.

FFPE was first described more than 100 years ago, and most hospitals still use this method today. But there has never been a set of standardized guidelines for processing FFPE tissue samples taken from patients to preserve tissue histology, let alone biomolecules. And although it works well for histology, the lack of standardized guidelines seems to have hampered the use of FFPE samples in molecular analyses. This may soon change, as pathologists are working towards standardizing FFPE sample preparation, and companies and researchers are developing the technology needed to isolate biomolecules and tap into the vast treasure

chest of archived samples.

Although FFPE tissue preparation is simple in theory, many problems associated with downstream molecular applications — such as PCR or microarray analysis — can arise. “This is all about the fact that there has been no attention paid to uniformity of preparation,” says David Rimm, a pathologist at Yale University. Between hospitals the time to tissue fixation and even the method of fixation can vary dramatically.

“I would say that the biggest issue is time from ligation of circulation to fixation,” says Rimm. During this period of ischaemia, molecular changes occur that cause problems in obtaining biomolecules. “Phosphorylation is very sensitive to ischaemic times. There seems to be promiscuous phosphatases in the cell that knock phosphates off tyrosines during this period,” says Rimm. DNA and RNA can also suffer damage before fixation, with



David Rimm: time is key in tissue analysis.

enzymes degrading and modifying both.

It seems obvious that rapid fixation is the answer, but no simple solution is in sight. Rimm says that for a researcher interested in only DNA or RNA, rapid fixation using quick freezing methods, instead of chemical fixation, is probably best. But if you are interested in proteins, freezing is problematic as the subsequent thawing process tends to break up and denature proteins and, like ischaemia,

can lead to modifications.

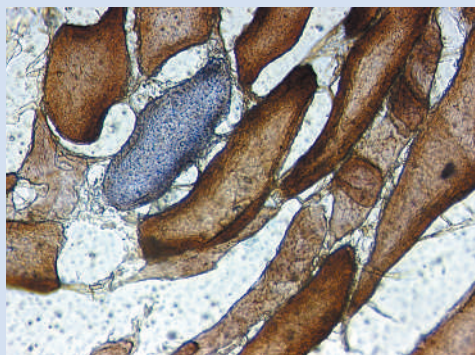
Then there are pathologists who think that standardizing the time to fixation will prove to be a difficult task. “Standardization would be great, but I don’t know how realistic it is across institutions or even within institutions,” says Christine Iacobuzio-Donahue, a pathologist at Johns Hopkins University School of Medicine in Baltimore, Maryland. She believes that the best place to tackle the problems caused by ischaemia is actually downstream of fixation,

D. RIMM

FROZEN IN TIME

Analysing proteins from tissue samples that have been formalin-fixed and paraffin-embedded (FFPE) can provide critical information about how cells function before fixation. “When you formalin-fix the protein, you fix it in time; it is not going anywhere,” says Peter Tunon, vice-president of sales and marketing at Expression Pathology in Gaithersburg, Maryland. Researchers are taking advantage of this fact to explore the protein world in FFPE samples.

Expression Pathology was founded in 2001 by researchers from the US National Cancer Institute and the company Life Technologies (now Invitrogen) who had experience in studying gene expression in tissue and histology. “The company was founded on the fact that examining protein expression is crucial to understanding what is happening in cells,” says Tunon. To this end, Expression Pathology has worked



Caught on camera: a section of human muscle after dual histochemistry.

on ways of extracting and isolating proteins from FFPE tissue samples for analysis by mass spectrometry or reverse-phase dot blot arrays. The company developed a technology that integrates extraction of total proteins from FFPE samples with tryptic digestion so that finished samples are ready for mass spectrometry. To simplify the system further,

all reagents are completely compatible with mass-spectrometry instrumentation, says Tunon, making it a good starting point for broad-based screening of proteins in FFPE samples. Other companies also think that protein isolation from

FFPE tissue samples will provide valuable information. QIAGEN and EMD Chemicals, both based in San Diego, California, now offer systems that chemically reverse formalin crosslinking to isolate full-length proteins for applications such as western blotting and protein arrays.

Surprisingly, post-translational modifications (PTMs) such as

phosphorylation and acetylation, can also be observed when examining proteins isolated from FFPE samples. PTMs can be critical to the role of a protein in the cell, changing the function or localization. “We do see post-translational modifications that are preserved on the peptides, and they seem to be present in ratios that are similar to those we get when using fresh frozen tissues,” Tunon says of work done by Expression Pathology.

Even much older archival samples of FFPE tissues do not seem to pose a problem for protein extraction. “We have worked with samples that are more than 15 years old and found no difference in the profiles compared to samples that were just a few weeks old,” says Tunon. He is quick to add that he believes even older samples could be examined, but Expression Pathology has yet to test this idea.

LEICA MICROSYSTEMS

N.B.

after the biomolecules have been isolated. "Once you have the samples and have extracted the biomolecules you are interested in, then you can perform quality control to determine if all your samples are similar," she says.

Even if the optimal time to fixation were found and standardized, this is only part of the problem; researchers still need to identify the best fixative to preserve both histology and biomolecules. Formalin, the most widely used fixative, provides good preservation of tissue histology but can cause problems for researchers interested in downstream molecular applications. "Formalin tends to crosslink the tissue to such an extent that it is hard to get fragments of DNA that are longer than, maybe, a hundred or two hundred base pairs," says Rimm. Formalin fixation causes the crosslinking of DNA as well as RNA and proteins, although proteins seem to fair better following formalin treatment (see 'Frozen in time').

Preservation society

Although tissue preservation has its problems, the good news is that these are now being dealt with. Governmental agencies, such as the US Office of Biorepositories and Biospecimen Research at the National Cancer Institute (NCI) in Bethesda, Maryland, are starting to tackle the difficult issue of biospecimen standardization. In June, the office released a guide to the NCI's best practices for biospecimen resources, detailing technical guidelines for NCI-supported biospecimen collection and storage. And researchers and companies are



The DASL assay can be processed on two different platforms: the Array Matrix (left) or the BeadChip.

now creating methods to use the degraded and modified biomolecules obtained from FFPE samples for molecular analysis.

Isolation of RNA or DNA from FFPE tissue samples can be accomplished using a number of methods or kits. The main problem is that, almost without fail, the RNA or DNA isolated is degraded and chemically modified. But because researchers want to tap into the vast archives of FFPE tissues for global expression analysis and biomarker discovery, this is spurring companies to address the issue.

Microarray analysis has proven a valuable tool for understanding global gene-expression patterns. But using the degraded messenger RNA obtained from FFPE tissues for such analysis is problematic. "The results on standard microarrays are currently unsatisfactory," says Shawn Baker, scientific product manager for gene expression at Illumina in San Diego, California. For this reason, Illumina offers a gene-expression application that can study RNA

extracted from FFPE tissue samples. Called DASL, short for cDNA-mediated annealing, selection, extension and ligation assay, this system amplifies the mRNAs from FFPE-extracted samples, but unlike other amplification systems it is not 3' biased, says Baker. "We use a combination of random and oligo dT primers to generate the complementary DNA, which means that even with degraded RNA it still amplifies quite well and produces good, consistent profiles." Following the first amplification, the DASL system uses two gene-specific probes to amplify the cDNA. The resulting cDNA can be hybridized to a DASL-specific array. "The DASL assay is multiplexed up to 1,536 genes. So you get a tremendous boost in the overall throughput," says Baker.

Although unable to survey as many genes as standard microarrays, Baker says that using the DASL system, researchers have been able to profile FFPE tissue samples that are up to 30 years old — demonstrating the potential to

THE CUTTING EDGE

Researchers have the opportunity to play surgeon — slicing and dissecting out specific sections of tissue or even cell populations — with laser-capture microdissection (LCM). This level of analysis might seem to be difficult when applied to formalin-fixed and paraffin-embedded (FFPE) tissue samples, but many companies are now offering easy and quick LCM solutions.

Leica Microsystems of Wetzlar, Germany, offers the LMD6000 LCM system, which uses an upright microscope for dissection and capture. Christoph Horlemann, the company's product manager for the LMD6000, says this is possible because the transport mechanism for capture is based on gravity, unlike other LCM systems on the market.

'Transport mechanism' refers to the method for delivering a dissected tissue sample from

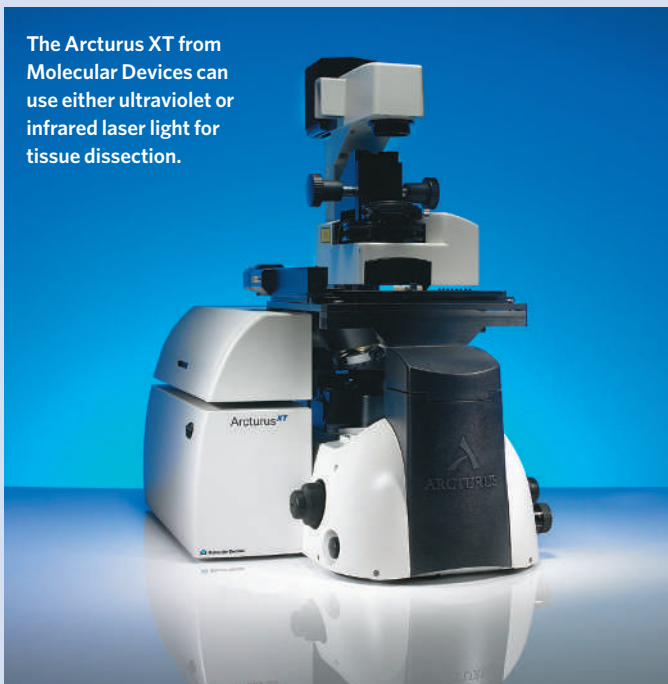
a slide to a collection vessel.

Other approaches to this process include that of PALM Microlaser Technologies in Bernried, Germany, which uses a 'pressure catapult' to send sections into tubes from the LCM instrument, and the CapSure system from Molecular Devices in Sunnyvale, California, which uses the laser to extend a polymer onto the tissue sample for capture.

Leica Microsystems' transport mechanism uses slides with a foil covering. The tissue is attached to the foil covering but not the slide, so researchers cut the tissue and the foil together and the sample simply falls by gravity into the collection tube below, explains Horlemann. But even this foil coating might not be needed in the future.

Both Leica and PALM are working with Expression Pathology of Gaithersburg, Maryland, on what may be

The Arcturus XT from Molecular Devices can use either ultraviolet or infrared laser light for tissue dissection.



the next generation of LCM slides. Called Director, these

glass slides are based on laser-induced forward transfer (LIFT),

examine the numerous collections of archived FFPE samples. Illumina is now working to increase the number of genes analysed for each DASL assay.

NuGEN Technologies in San Carlos, California, is another company developing methods to use RNA extracted from FFPE tissue samples for gene-expression analysis. NuGEN specializes in working with very small amounts of RNA or difficult-to-use RNA, such as RNA extracted from whole blood or very degraded sources. This is commonly seen with clinical samples, but is most significant in the case of FFPE, says Gianfranco de Feo, senior director of customer solutions at NuGEN. Although NuGEN did not start off looking at RNA from FFPE tissue samples, it was the next logical step. "We have had a product on the market for over a year now that allows users to work with very degraded RNA in very limited amounts, down to 500 picograms. We built on that technology to create kits for the much more degraded RNA that comes from FFPE samples," says de Feo.

At the core of NuGEN's technology is its amplification and labelling system, which has been optimized to work with Affymetrix 3' microarrays. The system relies on a combination of random hexamers, similar to that of Illumina, augmented with oligo (dT) primers to convert mRNA into cDNA in a linear amplification process. The inclusion of oligo (dT) primers was essential because the Affymetrix arrays probe the 3' ends of transcripts. But NuGEN hopes to have labelling and hybridization protocols and products for other microarray platforms

available before the end of the year.

To determine how well these degraded RNA samples from FFPE tissue will work on microarrays, NuGEN developed a tool using quantitative real-time PCR (qPCR) assay. It turns out that the results from this assay correlate very well with the overall results of microarray analysis, says de Feo, allowing researchers to decide whether the data that could be obtained from the array will be of sufficient quality to continue. And this is critical information, as in some cases less than 50% of the transcripts on the array may hybridize with the amplified RNA.

Think small

Asuragen in Austin, Texas, is a new company on the commercial block, working to understand and characterize the biological role of small RNAs. Although founded only a year ago, Asuragen's RNA roots go much deeper. Asuragen is a spin-off of Ambion, a company that worked in the field of molecular biology with a focus on RNA for nearly 17 years. "At Ambion, we developed the first kits and technologies for characterizing small RNAs," says Gary Latham, associate director of technology development at Asuragen. "And when microRNAs emerged as a new class of regulatory RNAs in humans, we were sitting in an excellent position to explore this area of 'biological dark matter'." In March 2006, Ambion was sold to Applied Biosystems for US\$273 million and a portion of those proceeds were used to fund Asuragen.

Asuragen has concentrated its efforts on the



Leica Microsystems offers the LMD6000 for laser capture microdissection applications.

diagnostic and therapeutic opportunities of microRNAs (miRNAs). "As they are smaller, miRNAs tend to survive the more tortuous conditions of FFPE tissue processing better than mRNAs," says Latham, which makes

a non-contact microdissection technique that uses a thin energy-transfer coating that replaces plastic films or adhesives. The technology was co-developed by researchers at the US Naval Research Laboratory in Washington DC and scientists from Expression Pathology.

Laser energy is transferred to the transfer layer and the layer is vaporized. The laser energy is then converted into kinetic energy, and the selected feature is shot instantly into a collection tube.

LIFT works equally well whether the tissue is collected up (PALM) or down (Leica) as the energy is sufficient to propel the tissue into the collection tube either way. As the transfer layer completely absorbs the laser energy, the biomolecules in the sample are not affected. The use of glass with this coating has other advantages as well.

"Glass slides are quite useful as

you can also perform fluorescence and contrast applications without any interfering foil," says Horlemann.

Molecular Devices, a division of MDS Analytical Technologies, acquired Arcturus in 2006 and is now providing both the Veritas and the Arcturus XT LCM systems. Whereas most LCM systems use an ultraviolet laser for cutting samples, the systems offered by Molecular Devices can have either ultraviolet or the standard infrared laser options.

"The infrared laser is ideal for small areas where a user is looking to pick up only a cell or two," says Steven Blakely, product manager for the Arcturus LCM systems. "Infrared allows for a gentle collection of cells."

The use of infrared laser light is critical to the CapSure system because the dye found in the CapSure polymer is activated and becomes adherent with infrared

light. Blakely says the CapSure system is particularly useful for FFPE samples that are already mounted on glass slides because the transport mechanism involves only the polymer adhering to the tissue and so can work with any glass slide.

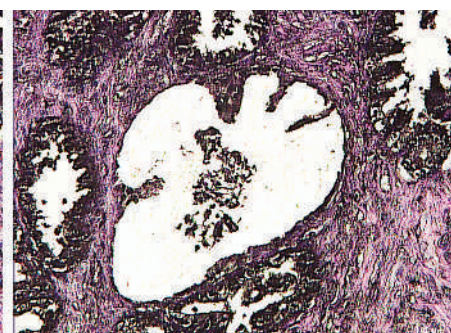
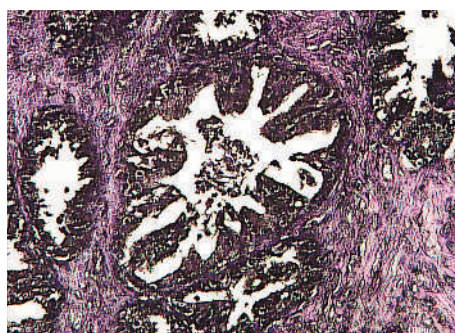
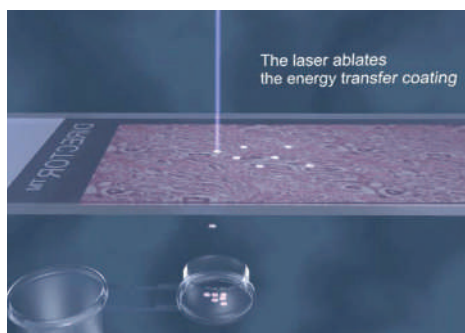
Cutting tissue samples with ultraviolet light offers advantages too, such as dissecting thicker samples. Leica has worked to optimize objectives for its LCM systems with high-energy ultraviolet-light transmission, allowing more power to come in direct contact with the tissue sample. Horlemann says this provides faster cutting of the thicker tissue samples while using less power.

The future of LCM might just lie in automation and increases in throughput. "Now, researchers want to do faster, automated microdissection for proteomics," says Horlemann. But before

anything else, some hardware and software issues needed to be resolved.

Any automated software package has to control all microdissection steps including focusing the microscope, recognizing cells of interest, focusing the laser and defining the area to cut — definitely not a simple task, but one that companies have worked on and made tremendous strides in recent years. Horlemann says that new advances are making automation easier every day. He points to the Director slides, which can allow for contrasting methods and fluorescence, making it easier when attempting to automatically define cells of interest as a step in the right direction.

Although simpler now, further advances will be required before LCM becomes as easy as pushing a button and walking away. N.B.



EXPRESSION PATHOLOGY

Direct action: Expression Pathology's Director slide (left) in action, and tissue samples before (middle) and after laser capture microdissection with the system.

degradation less of an issue. This means that downstream applications, such as microarrays, are better suited to miRNAs isolated from FFPE samples than mRNAs, says Latham. "We have found that there is tremendous diagnostic potential for microRNAs as biomarkers of disease states," he adds. This shows how companies are moving from traditional RNA analysis to non-traditional methods that use FFPE tissue samples.

Several other companies also offer technologies to explore RNA isolated from FFPE tissue samples. Panomics in Fremont, California, has developed a direct hybridization method using branched DNA technology. The system, called QuantiGene FFPE, is unique because it does not involve the linear amplification of RNA, as do many other systems. Instead, the use of branched DNA technology permits the direct measurement of RNA from the sample source. The Paradise System, developed by Arcturus Biosciences and now supplied by Molecular Devices of Sunnyvale, California, has been optimized for RNA extraction from FFPE samples, followed by a linear amplification step prior to use in qPCR or microarray applications.

Even though much headway has been made in the molecular analysis of FFPE tissue samples over the past few years, high-throughput solutions to examine the billions of archived FFPE tissue samples are still needed. But advances in technologies, including tissue microarrays (TMAs) and laser-capture

microdissection (see 'The cutting edge'), are signalling that high-throughput analysis might be around the corner.

In 1998, Juha Kononen and his colleagues described a tissue sampling method that produced regular-sized spots that could be densely packed on a microarray slide (J. Kononen *et al. Nature Med.* 4, 844–847; 1998). Using this methodology, archival FFPE tissues can be sampled onto TMAs, allowing researchers to examine numerous samples by techniques such as fluorescence *in situ* hybridization and immunohistochemistry on a single microscope slide. Nucleic acids and proteins can even be extracted from archival FFPE tissue TMAs. Best of all, the technology for constructing TMAs is readily available from commercial suppliers.

Spot the tissue

Beecher Instruments of Sun Prairie, Wisconsin, produces both manual and automated tissue microarrays. With the Manual Tissue Arrayer II, a block can be directly attached to a microtome for sectioning of arrays, whereas the automated arrayer, ATA-27, accepts nearly all tissue cassettes and can be adapted for either large arrays or variously shaped archival samples. The ATA-27 can also accommodate a full range of tissue spot sizes from 0.6 to 3 millimetres.

One concern regarding TMAs has been the representation of a whole tissue section by a single spot. "As the throughput is so much higher, I think that trumps the fact that it is not a whole section," says Rimm. But he cautions that care must be taken when constructing tissue microarrays. The optimal region to be sampled for the microarray is usually identified by a pathologist, but only a section or core of this region can be placed into the array. Problems can arise because the

region that cores are taken from can vary. A core taken from the leading edge of a tumour might provide different results from a core acquired from the middle. So design of the slide is crucial; spots should be separated and standards (spots of known tissues) used.

In the past, pathologists have evaluated and analysed TMAs following immunohistochemistry or other histological analyses. But that is now changing, and some companies are offering software specifically for automated TMA analysis.

Image bank

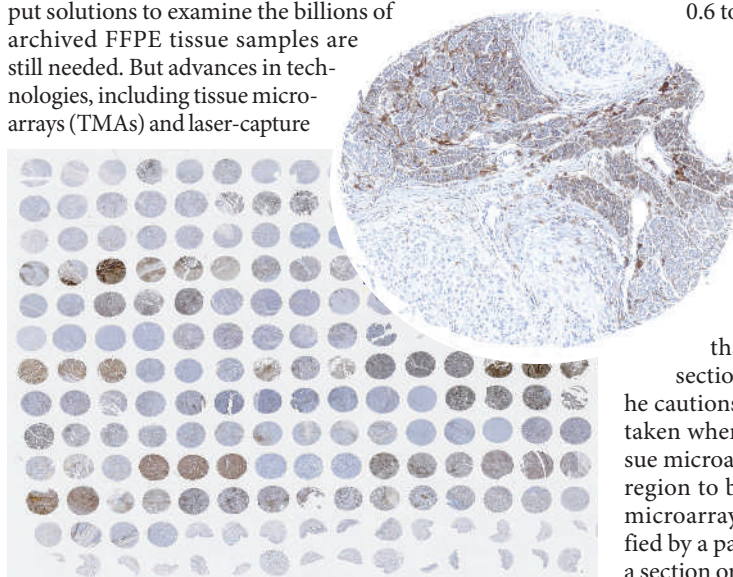
Bacus Laboratories in Lombard, Illinois, and recently acquired by Olympus, is one such company. It has focused its efforts on virtual microscopy — the digital imaging of microscope slides — which involves scanning an entire slide at very high resolution to acquire a large number of images for each slide. The images are stored in a database and the virtual slide can be reconstructed using software that puts the single images back together to form a whole.

It was virtual microscopy that led Bacus into the world of tissue microarrays. "Several years ago, there were no good scanning methods for tissue microarrays," says James Bacus, the company's president, "but we had the ability to scan entire slides." Using virtual microscopy as a starting point, Bacus has developed several programs for quantitative, automated TMA analysis including TMAScore and IHCscore.

Several other companies also provide automated TMA analysis solutions. Beecher Instruments has the TMAx software package that relies on morphometric processing algorithms and classification rules to automatically interpret TMAs. And HistoRX in New Haven, Connecticut, offers the AQUA automated quantitative pathology system for biomarker discovery and validation from tissue samples.

With TMAs moving FFPE analysis onto the fast track of high-throughput analysis, all the pieces are falling into place for researchers to begin exploring archival FFPE tissue samples. The coming years could prove very interesting indeed as the molecular secrets of these FFPE samples are finally unlocked.

Nathan Blow is the technology editor for Nature and Nature Methods.



Tissue microarrays containing single spots of tissue (inset) provide a high-throughput histology solution.

Tissue issues

Millions of tissue samples have been collected and archived, but researchers wanting to explore them at the molecular level have found it tough going. **Nathan Blow** investigates the issues.

According to experts, there are more than a billion tissue samples archived in hospitals and tissue banks around the world, most of them formalin-fixed and paraffin-embedded (FFPE). Today, these samples present both an incredible opportunity and a huge challenge to researchers. FFPE tissue samples have been extensively annotated and well preserved, allowing detailed study of the progression of diseases such as cancer. But due to the method of preservation, obtaining biomolecules from these samples is proving difficult, to say the least.

FFPE was first described more than 100 years ago, and most hospitals still use this method today. But there has never been a set of standardized guidelines for processing FFPE tissue samples taken from patients to preserve tissue histology, let alone biomolecules. And although it works well for histology, the lack of standardized guidelines seems to have hampered the use of FFPE samples in molecular analyses. This may soon change, as pathologists are working towards standardizing FFPE sample preparation, and companies and researchers are developing the technology needed to isolate biomolecules and tap into the vast treasure

chest of archived samples.

Although FFPE tissue preparation is simple in theory, many problems associated with downstream molecular applications — such as PCR or microarray analysis — can arise. “This is all about the fact that there has been no attention paid to uniformity of preparation,” says David Rimm, a pathologist at Yale University. Between hospitals the time to tissue fixation and even the method of fixation can vary dramatically.

“I would say that the biggest issue is time from ligation of circulation to fixation,” says Rimm. During this period of ischaemia, molecular changes occur that cause problems in obtaining biomolecules. “Phosphorylation is very sensitive to ischaemic times. There seems to be promiscuous phosphatases in the cell that knock phosphates off tyrosines during this period,” says Rimm. DNA and RNA can also suffer damage before fixation, with



David Rimm: time is key in tissue analysis.

enzymes degrading and modifying both.

It seems obvious that rapid fixation is the answer, but no simple solution is in sight. Rimm says that for a researcher interested in only DNA or RNA, rapid fixation using quick freezing methods, instead of chemical fixation, is probably best. But if you are interested in proteins, freezing is problematic as the subsequent thawing process tends to break up and denature proteins and, like ischaemia,

can lead to modifications.

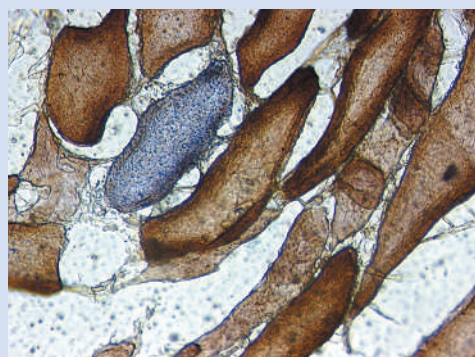
Then there are pathologists who think that standardizing the time to fixation will prove to be a difficult task. “Standardization would be great, but I don’t know how realistic it is across institutions or even within institutions,” says Christine Iacobuzio-Donahue, a pathologist at Johns Hopkins University School of Medicine in Baltimore, Maryland. She believes that the best place to tackle the problems caused by ischaemia is actually downstream of fixation,

D. RIMM

FROZEN IN TIME

Analysing proteins from tissue samples that have been formalin-fixed and paraffin-embedded (FFPE) can provide critical information about how cells function before fixation. “When you formalin-fix the protein, you fix it in time; it is not going anywhere,” says Peter Tunon, vice-president of sales and marketing at Expression Pathology in Gaithersburg, Maryland. Researchers are taking advantage of this fact to explore the protein world in FFPE samples.

Expression Pathology was founded in 2001 by researchers from the US National Cancer Institute and the company Life Technologies (now Invitrogen) who had experience in studying gene expression in tissue and histology. “The company was founded on the fact that examining protein expression is crucial to understanding what is happening in cells,” says Tunon. To this end, Expression Pathology has worked



Caught on camera: a section of human muscle after dual histochemistry.

on ways of extracting and isolating proteins from FFPE tissue samples for analysis by mass spectrometry or reverse-phase dot blot arrays. The company developed a technology that integrates extraction of total proteins from FFPE samples with tryptic digestion so that finished samples are ready for mass spectrometry. To simplify the system further,

all reagents are completely compatible with mass-spectrometry instrumentation, says Tunon, making it a good starting point for broad-based screening of proteins in FFPE samples. Other companies also think that protein isolation from

FFPE tissue samples will provide valuable information. QIAGEN and EMD Chemicals, both based in San Diego, California, now offer systems that chemically reverse formalin crosslinking to isolate full-length proteins for applications such as western blotting and protein arrays.

Surprisingly, post-translational modifications (PTMs) such as

phosphorylation and acetylation, can also be observed when examining proteins isolated from FFPE samples. PTMs can be critical to the role of a protein in the cell, changing the function or localization. “We do see post-translational modifications that are preserved on the peptides, and they seem to be present in ratios that are similar to those we get when using fresh frozen tissues,” Tunon says of work done by Expression Pathology.

Even much older archival samples of FFPE tissues do not seem to pose a problem for protein extraction. “We have worked with samples that are more than 15 years old and found no difference in the profiles compared to samples that were just a few weeks old,” says Tunon. He is quick to add that he believes even older samples could be examined, but Expression Pathology has yet to test this idea.

LEICA MICROSYSTEMS

N.B.

after the biomolecules have been isolated. "Once you have the samples and have extracted the biomolecules you are interested in, then you can perform quality control to determine if all your samples are similar," she says.

Even if the optimal time to fixation were found and standardized, this is only part of the problem; researchers still need to identify the best fixative to preserve both histology and biomolecules. Formalin, the most widely used fixative, provides good preservation of tissue histology but can cause problems for researchers interested in downstream molecular applications. "Formalin tends to crosslink the tissue to such an extent that it is hard to get fragments of DNA that are longer than, maybe, a hundred or two hundred base pairs," says Rimm. Formalin fixation causes the crosslinking of DNA as well as RNA and proteins, although proteins seem to fair better following formalin treatment (see 'Frozen in time').

Preservation society

Although tissue preservation has its problems, the good news is that these are now being dealt with. Governmental agencies, such as the US Office of Biorepositories and Biospecimen Research at the National Cancer Institute (NCI) in Bethesda, Maryland, are starting to tackle the difficult issue of biospecimen standardization. In June, the office released a guide to the NCI's best practices for biospecimen resources, detailing technical guidelines for NCI-supported biospecimen collection and storage. And researchers and companies are



The DASL assay can be processed on two different platforms: the Array Matrix (left) or the BeadChip.

now creating methods to use the degraded and modified biomolecules obtained from FFPE samples for molecular analysis.

Isolation of RNA or DNA from FFPE tissue samples can be accomplished using a number of methods or kits. The main problem is that, almost without fail, the RNA or DNA isolated is degraded and chemically modified. But because researchers want to tap into the vast archives of FFPE tissues for global expression analysis and biomarker discovery, this is spurring companies to address the issue.

Microarray analysis has proven a valuable tool for understanding global gene-expression patterns. But using the degraded messenger RNA obtained from FFPE tissues for such analysis is problematic. "The results on standard microarrays are currently unsatisfactory," says Shawn Baker, scientific product manager for gene expression at Illumina in San Diego, California. For this reason, Illumina offers a gene-expression application that can study RNA

extracted from FFPE tissue samples. Called DASL, short for cDNA-mediated annealing, selection, extension and ligation assay, this system amplifies the mRNAs from FFPE-extracted samples, but unlike other amplification systems it is not 3' biased, says Baker. "We use a combination of random and oligo dT primers to generate the complementary DNA, which means that even with degraded RNA it still amplifies quite well and produces good, consistent profiles." Following the first amplification, the DASL system uses two gene-specific probes to amplify the cDNA. The resulting cDNA can be hybridized to a DASL-specific array. "The DASL assay is multiplexed up to 1,536 genes. So you get a tremendous boost in the overall throughput," says Baker.

Although unable to survey as many genes as standard microarrays, Baker says that using the DASL system, researchers have been able to profile FFPE tissue samples that are up to 30 years old — demonstrating the potential to

THE CUTTING EDGE

Researchers have the opportunity to play surgeon — slicing and dissecting out specific sections of tissue or even cell populations — with laser-capture microdissection (LCM). This level of analysis might seem to be difficult when applied to formalin-fixed and paraffin-embedded (FFPE) tissue samples, but many companies are now offering easy and quick LCM solutions.

Leica Microsystems of Wetzlar, Germany, offers the LMD6000 LCM system, which uses an upright microscope for dissection and capture. Christoph Horlemann, the company's product manager for the LMD6000, says this is possible because the transport mechanism for capture is based on gravity, unlike other LCM systems on the market.

'Transport mechanism' refers to the method for delivering a dissected tissue sample from

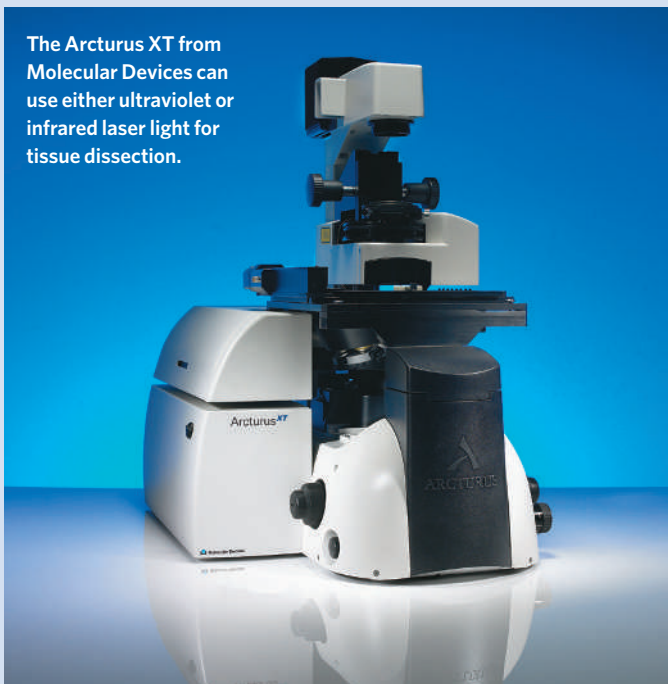
a slide to a collection vessel.

Other approaches to this process include that of PALM Microlaser Technologies in Bernried, Germany, which uses a 'pressure catapult' to send sections into tubes from the LCM instrument, and the CapSure system from Molecular Devices in Sunnyvale, California, which uses the laser to extend a polymer onto the tissue sample for capture.

Leica Microsystems' transport mechanism uses slides with a foil covering. The tissue is attached to the foil covering but not the slide, so researchers cut the tissue and the foil together and the sample simply falls by gravity into the collection tube below, explains Horlemann. But even this foil coating might not be needed in the future.

Both Leica and PALM are working with Expression Pathology of Gaithersburg, Maryland, on what may be

The Arcturus XT from Molecular Devices can use either ultraviolet or infrared laser light for tissue dissection.



the next generation of LCM slides. Called Director, these

glass slides are based on laser-induced forward transfer (LIFT),

examine the numerous collections of archived FFPE samples. Illumina is now working to increase the number of genes analysed for each DASL assay.

NuGEN Technologies in San Carlos, California, is another company developing methods to use RNA extracted from FFPE tissue samples for gene-expression analysis. NuGEN specializes in working with very small amounts of RNA or difficult-to-use RNA, such as RNA extracted from whole blood or very degraded sources. This is commonly seen with clinical samples, but is most significant in the case of FFPE, says Gianfranco de Feo, senior director of customer solutions at NuGEN. Although NuGEN did not start off looking at RNA from FFPE tissue samples, it was the next logical step. "We have had a product on the market for over a year now that allows users to work with very degraded RNA in very limited amounts, down to 500 picograms. We built on that technology to create kits for the much more degraded RNA that comes from FFPE samples," says de Feo.

At the core of NuGEN's technology is its amplification and labelling system, which has been optimized to work with Affymetrix 3' microarrays. The system relies on a combination of random hexamers, similar to that of Illumina, augmented with oligo (dT) primers to convert mRNA into cDNA in a linear amplification process. The inclusion of oligo (dT) primers was essential because the Affymetrix arrays probe the 3' ends of transcripts. But NuGEN hopes to have labelling and hybridization protocols and products for other microarray platforms

available before the end of the year.

To determine how well these degraded RNA samples from FFPE tissue will work on microarrays, NuGEN developed a tool using quantitative real-time PCR (qPCR) assay. It turns out that the results from this assay correlate very well with the overall results of microarray analysis, says de Feo, allowing researchers to decide whether the data that could be obtained from the array will be of sufficient quality to continue. And this is critical information, as in some cases less than 50% of the transcripts on the array may hybridize with the amplified RNA.

Think small

Asuragen in Austin, Texas, is a new company on the commercial block, working to understand and characterize the biological role of small RNAs. Although founded only a year ago, Asuragen's RNA roots go much deeper. Asuragen is a spin-off of Ambion, a company that worked in the field of molecular biology with a focus on RNA for nearly 17 years. "At Ambion, we developed the first kits and technologies for characterizing small RNAs," says Gary Latham, associate director of technology development at Asuragen. "And when microRNAs emerged as a new class of regulatory RNAs in humans, we were sitting in an excellent position to explore this area of 'biological dark matter'." In March 2006, Ambion was sold to Applied Biosystems for US\$273 million and a portion of those proceeds were used to fund Asuragen.

Asuragen has concentrated its efforts on the



Leica Microsystems offers the LMD6000 for laser capture microdissection applications.

diagnostic and therapeutic opportunities of microRNAs (miRNAs). "As they are smaller, miRNAs tend to survive the more tortuous conditions of FFPE tissue processing better than mRNAs," says Latham, which makes

a non-contact microdissection technique that uses a thin energy-transfer coating that replaces plastic films or adhesives. The technology was co-developed by researchers at the US Naval Research Laboratory in Washington DC and scientists from Expression Pathology.

Laser energy is transferred to the transfer layer and the layer is vaporized. The laser energy is then converted into kinetic energy, and the selected feature is shot instantly into a collection tube.

LIFT works equally well whether the tissue is collected up (PALM) or down (Leica) as the energy is sufficient to propel the tissue into the collection tube either way. As the transfer layer completely absorbs the laser energy, the biomolecules in the sample are not affected. The use of glass with this coating has other advantages as well.

"Glass slides are quite useful as

you can also perform fluorescence and contrast applications without any interfering foil," says Horlemann.

Molecular Devices, a division of MDS Analytical Technologies, acquired Arcturus in 2006 and is now providing both the Veritas and the Arcturus XT LCM systems. Whereas most LCM systems use an ultraviolet laser for cutting samples, the systems offered by Molecular Devices can have either ultraviolet or the standard infrared laser options.

"The infrared laser is ideal for small areas where a user is looking to pick up only a cell or two," says Steven Blakely, product manager for the Arcturus LCM systems. "Infrared allows for a gentle collection of cells."

The use of infrared laser light is critical to the CapSure system because the dye found in the CapSure polymer is activated and becomes adherent with infrared

light. Blakely says the CapSure system is particularly useful for FFPE samples that are already mounted on glass slides because the transport mechanism involves only the polymer adhering to the tissue and so can work with any glass slide.

Cutting tissue samples with ultraviolet light offers advantages too, such as dissecting thicker samples. Leica has worked to optimize objectives for its LCM systems with high-energy ultraviolet-light transmission, allowing more power to come in direct contact with the tissue sample. Horlemann says this provides faster cutting of the thicker tissue samples while using less power.

The future of LCM might just lie in automation and increases in throughput. "Now, researchers want to do faster, automated microdissection for proteomics," says Horlemann. But before

anything else, some hardware and software issues needed to be resolved.

Any automated software package has to control all microdissection steps including focusing the microscope, recognizing cells of interest, focusing the laser and defining the area to cut — definitely not a simple task, but one that companies have worked on and made tremendous strides in recent years. Horlemann says that new advances are making automation easier every day. He points to the Director slides, which can allow for contrasting methods and fluorescence, making it easier when attempting to automatically define cells of interest as a step in the right direction.

Although simpler now, further advances will be required before LCM becomes as easy as pushing a button and walking away. N.B.

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
Gene expression analysis			
Affymetrix	SNP genotyping; microarray analysis	Santa Clara, California	www.affymetrix.com
Aglient	Microarray; mass spectrometry; lab-on-a-chip methods	Wilmington, Delaware	www.aglient.com
Applied Biosystems	Real-time PCR; SNP genotyping; microarrays; microRNA analysis	Foster City, California	www.appliedbiosystems.com
Asuragen	miRNA-based applications and therapeutics; cancer detection; pharmacogenomics	Austin, Texas	www.asuragen.com
BioDiscovery	Microarray analysis software provider	El Segundo, California	www.biodiscovery.com
Bio-Rad Laboratories	Real-time PCR instruments and accessories; microarray accessories	Hercules, California	www.bio-rad.com
Ciphergen	Biomarker identification for drug discovery and molecular imaging of targets	Fremont, California	www.ciphergen.com
Chemicon	PCR kits	Temecula, California	www.chemicon.com
EMD Chemicals	Kits for extraction of proteins from FFPE tissue samples	San Diego, California	www.emdbiosciences.com
Eppendorf	DualChip DNA microarrays	Hamburg, Germany	www.eppendorf.com
Expression Analysis	Microarray-based genomics services for clinical trials	Durham, North Carolina	www.expressionanalysis.com
GE Healthcare	Microarrays; sequencing; SNP genotyping	Piscataway, New Jersey	www.gehealthcare.com
Genomic Solutions	Assay assembly for SNP genotyping, sequencing, PCR; microarray printing	Ann Arbor, Michigan	www.genomesolutions.com
Idaho Technology	RapidCycler for real-time PCR	Lake City, Utah	www.idahotech.com
Illumina	SNP analysis; microarrays; DNA sequencing; DASL system	San Diego, California	www.illumina.com
Integrated DNA Technologies	Custom oligonucleotides for microarrays	Coralville, Iowa	www.idtdna.com
Invitrogen	Microarray accessories; array comparative genomic hybridization kits; RT PCR; gene-expression analysis services	San Diego, California	www.invitrogen.com
NuGEN	RNA isolation and analysis	San Carlos, California	www.nugeninc.com
Operon Biotechnologies	Preprinted oligonucleotide arrays and probe sets	Huntsville, Alabama	www.operon.com
Orchid Cellmark	SNP analysis for degraded samples; forensic analysis	Princeton, New Jersey	www.orchidbio.com
Oxford Gene Technology	Oligonucleotide arrays for various applications	Oxford, UK	www.ogt.co.uk
Panomics	Quantitative gene-expression measurements using QuantiGene system; protein and DNA arrays; cytokine arrays	Fremont, California	www.panomics.com
Primer Biosystems	Multiplex RT PCR; microarray	Mansfield, Massachusetts	www.primerbio.com
QIAGEN	RT PCR and real-time PCR systems; epigenetics; laser microdissection accessories; RNA-extraction systems	Germantown, Maryland	www.qiagen.com
SuperArray	Real-time PCR; focused microarray; RNAi; gene-expression analysis services	Fredrick, Maryland	www.superarray.com
TeleChem International	Microarray robots, printers and scanners; microarray services; microarray accessories such as glass slides	Sunnyvale, California	www.arrayit.com
Laser microdissection systems and accessories			
Bücher Biotech	Laser-capture microdissection systems	Basel, Switzerland	www.bucher.ch
Carl Zeiss	PALM laser microdissector; PALM capture system	Berlin, Germany	www.zeiss.de
Erie Scientific Company	Microscope slides for use in laser-capture microdissection	Portsmouth, New Hampshire	www.eriesci.com
Expression Pathology	Slides for laser-capture microdissection applications	Gaithersburg, Maryland	www.expressionpathology.com
Leica Microsystems	Microscopes, microscopy software, laser-capture microdissection	Wetzlar, Germany	www.leica-microsystems.com
Molecular Devices	Laser-capture microdissection system; CapSure capture technology	Sunnyvale, California	www.molecular-device.com
Molecular Machines and Industries	Laser microdissection; micromanipulations	Glattbrugg, Switzerland	www.molecular-machines.com
PALM Microlaser Technologies	Non-contact laser-based micromanipulation	Bernried, Germany	www.palm-microlaser.com
Tissue microarray products and suppliers			
AMS Biotechnology	Frozen and paraffin-based tissue microarrays	Abingdon, UK	www.amsbio.com
Aperio	Digital tissue microarray analysis	Vista, California	www.aperio.com
Bacus Laboratories	Virtual microscopy; automated analysis of tissue microarrays; BLISS virtual scanner system	Lombard, Illinois	www.baculabs.com
Beecher Instruments	Manual and automated tissue microarrays; accessories for microarray construction and fabrication	Sun Prairie, Wisconsin	www.beecherinstruments.com
Biochain	Frozen and paraffin-based tissue microarrays; antibodies; DNA extraction kits	Hayward, California	www.biochain.com
Bioimage	Digital pathology; software for the analysis of tissue microarrays	Cupertino, California	www.bioimage.com
Cybrdi	Tissue microarrays; clinical specimens; immunohistochemistry services	Frederick, Maryland	www.cybrdi.com
Cytomix	Tissue biorepository; RT PCR cDNA panels; tissue sections, tissue microarrays	Lexington, Massachusetts	www.cytomix.com
HistoGeneX	Services for the detection and localization of cell-specific gene expression on paraffin-embedded tissue samples	Edegem, Belgium	www.histogenex.com
HistoRx	Automated analysis of immunohistochemistry	New Haven, Connecticut	www.historx.com
Imgenex	Histo-arrays from human and animal tissue sources; tissue sections; antibodies	San Diego, California	www.imgenex.com

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
Pathology Devices	Semi-automated tissue microarrayer and accessories	Westminster, Maryland	www.pathologydevices.com
Petagen	Accumax tissue microarrays	Seoul, Korea	www.petagen.com
Premier Biosoft International	Tissue-microarray analysis software for biomarker relatedness	Palo Alto, California	www.premierbiosoft.com
Protein Biotechnologies	Tissue-microarray supplier; human tissue cancer lysates	Ramona, California	www.proteinbiotechnologies.com
SuperBiochips	Human and animal tissue microarrays; immunohistochemistry services	Seoul, Korea	www.tissue-array.com
Unitma	Tissue microarrays; tissue microarray fabrication systems	Seoul, Korea	www.unitma.com
US BioMax	Tissue microarrays; tissue banking protein arrays; qPCR	Ijamsville, Maryland	www.usbiomax.com
Zyagen	Tissue samples; tissue microarrays; total RNA; cDNA; tissue section services	San Diego, California	www.zyagen.com

Microscopy and imaging

Andor Technology	CCD and intensified CCD camera systems	Belfast, UK	www.andor-tech.com	
Applied Imaging	Developed Ariol high-throughput immunohistochemistry system	San Jose, California	www.aicorp.com	
Applied Cytometry Systems	Software and cytometry systems	Sheffield, UK	www.appliedcytometry.com	
Applied Scientific	Automated systems for microscopy and fluorescence	Eugene, Oregon	www.asiimaging.com	
Biomedical Photometric	Tissue imaging instrumentation; high-resolution whole tissue imaging	Waterloo, Canada	www.confocal.com	
Bioptonics	Optical projection tomography scanning	Edinburgh, UK	www.bioptonics.com	
Bitplane	Imaris image analysis software	Zurich, Switzerland	www.bitplane.com	
Chroma Technology	Optical filter sets	Brattleboro, Vermont	www.chroma.com	
Clemex	Image analysis software	Longueuil, Quebec	www.clemex.com	
Fujifilm	Imaging systems for array analysis	Stamford, Connecticut	www.fujimed.com	●
Hamamatsu Photonics	Cameras for microscopy, light sources	Hamamatsu City, Japan	www.hamamatsu.com	
Kodak	Image analysis software; imaging systems	Rochester, New York	www.kodak.com	
Micro Video Instruments	Microscopes and accessories	Avon, Massachusetts	www.mvi-inc.com	
Nikon Instruments	Microscopes, microscopy analysis software and accessories	Melville, New York	www.nikoninstruments.com	●
Olympus	Microscopes, analysis software, accessories	Center Valley, Pennsylvania	www.olympus.com	
Omega Optical	Microscopes and filters	Brattleboro, Vermont	www.omegafilters.com	
Vision Biosystems	Automated immunohistochemistry systems and reagents	Mount Waverly, Australia	www.vision-bio.com	

General

Beckman Coulter	Tools and systems for molecular biology, genomics and proteomics	Fullerton, California	www.beckmancoulter.com	
BMG Labtech	Microplate readers and handling systems	Offenburg, Germany	www.bmg-labtechnologies.com	
Cambrex	Molecular and cell-biology reagents and tools	East Rutherford, New Jersey	www.cambrex.com	
Epicentre Biotechnologies	Enzymes for PCR and RT PCR; DNA and RNA purification	Madison, Wisconsin	www.epibio.com	
Hamilton Company	Robotics for life sciences	Reno, Nevada	www.hamiltoncomp.com	
Lonza	Molecular-biology reagents and systems	Basel, Switzerland	www.lonza.com	
Merck	Chemicals kits and reagents for molecular and cell-biology research	Darmstadt, Germany	www.merck.de	
Millipore	Automated equipment for genomics and proteomics	Billerica, Massachusetts	www.millipore.com	
MP Biomedicals	Reagents and chemicals for research	Aurora, Ohio	www.mpbio.com	
New Brunswick Scientific	Tissue culture; incubators; shakers	Edison, New Jersey	www.nbsc.com	
New England Biolabs	Molecular-biology reagents, kits and enzymes	Ipswich, Massachusetts	www.neb.com	
Sigma-Aldrich	PCR buffers and enzymes; catalogue chemical supplier	St Louis, Missouri	www.sigmaaldrich.com	●
Stoelting	Physiology research equipment, histology equipment	Wood Dale, Illinois	www.stoeltingco.com	
Takara Bio	Reagents, kits and services for genomics and molecular biology	Shiga, Japan	www.takara-bio.com	●
USB	Chemicals and reagents for molecular biology	Cleveland, Ohio	www.usweb.com	●

● see advertisement

naturejobs

JOBS OF THE WEEK

The drug industry has not been kind to its employees of late. With profits falling, pharmaceutical companies have been laying off staff, and thousands of scientists have felt the pinch. In January, Pfizer cut its work force by about 10%, a loss of 10,000 jobs. In July, AstraZeneca announced it would cut 7,600 jobs, again about 10% of its workforce. A few days later, Johnson & Johnson announced cuts of 4,800 jobs (about 4% of the workforce). The cuts affected employees across the United States and Europe. And last week, biotechnology firm Amgen announced lay-offs of between 2,200 and 2,600 people (12-14% of its workforce) — the company's first major lay-off.

Such 'realignments' have reverberating effects on the life-sciences job market and beyond. Along with scientists, marketing, sales and support staff have lost their jobs. Some biotechnology firms, which increasingly rely on drug-firm partners for revenue, have been forced to downsize as well.

Most drug companies blame dried-up drug pipelines. Blockbuster pills such as Viagra are increasingly rare finds, and the drugs that are far along in the development stages are increasingly hard to get approved for sale. High-profile tragedies, such as those related to Merck's Vioxx, haven't helped matters.

But there are still jobs out there. As we reported last year (see *Nature* **439**, 886-887; 2006), companies continue to recruit even in lean years. And the most recent numbers from the US Bureau of Labor Statistics suggest that employment for chemists and biologists in the US manufacturing sector for pharmaceutical and medicine is expected to grow by about 25% by 2014.

So scientists looking to industry for a livelihood need to do their homework. When and where were the company's last lay-offs? What are the company's bread-and-butter product lines? When do the patents run out? Do generic drugs threaten the company's bottom line? What else is in the pipeline?

Of course, there's no ignoring the most commonsense question of all — in the event of one of those inconvenient realignments, what's my plan B?

Gene Russo, acting editor of *Naturejobs*

CONTACTS

Acting Editor: Gene Russo

European Head Office, London

The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:

Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com

Natureevents:

Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:

Muriel Lestringuez (4994)

Southwest UK/RoW:

Nils Moeller (4953)

Scandinavia/Spain/Portugal/Italy:

Evelina Rubio-Hakansson (4973)

Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

North Germany/The Netherlands:

Reya Silao (4970)

South Germany/Austria:

Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:

Stephen Russell
To send materials use London
address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
Naturejobs web development:
Tom Hancock

Naturejobs online production:

Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com

Formic gender disorder

Divide and conquer.

Barrington J. Bayley

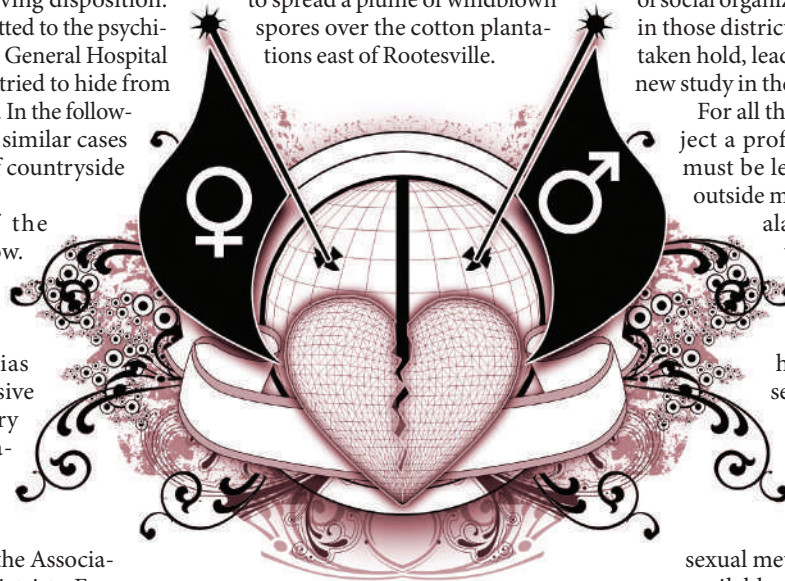
The waywardly named formic gender disorder (FGD) is the foremost of a growing number of ailments some physicians still refuse to recognize. *The Encyclopaedia of Psychocentric Dysfunctions*, compiled by Hortense Constantia, MD, in 2025, lists it alongside other disorders dismissed as imaginary. The first case to come to attention was that of Hattie Jacquiefs, a 45-year-old housewife of Rootesville, Louisiana, who in 2022 ran screaming from her house in terror of her husband, a man of known placid and loving disposition. Mrs Jacquiefs was admitted to the psychiatric ward of Rootesville General Hospital where she cowered and tried to hide from certain members of staff. In the following months a cluster of similar cases occurred in a swathe of countryside to the east of the town.

Determination of the aetiology has been slow. Initially investigation was psychiatric and the condition classified alongside other phobias and obsessive-compulsive disorders. Moves to carry out physical examinations were delayed by political objections, particularly on the part of a feminist faction in the Association of American Psychiatrists. Even now sufferers universally withhold consent to microbial studies. Only after some years was the disorder dignified with a diagnosis of somatic pathology as the view gradually began to prevail that it is a fungal infection of the nervous system. Relevant references are appended.

Federal authorities have now been forced to admit that an eastern suburb of Rootesville is home to the Canine and Feline Accidental Injuries Statistical Institute (CFAISI), a clandestine micropathology research laboratory funded by the Central Intelligence Agency. The laboratory had for 20 years been studying a nervous disease affecting the solitary stink ant, a denizen of the Western Cameroons. Occasionally the stink ant breathes in the spores of a fungus, which then grows inside its body, attacks its brain and controls its behaviour. Under the influence of the fungus the stink ant, for the first time in its life, climbs the stem of a plant, sinks its mandibles into it and clings there until it is consumed by the fun-

gus, having gained sufficient height from which to spread fresh spores.

A behaviour-controlling infectious disease has obvious political value. Formic gender disorder was named in honour of the African solitary stink ant by Doctors Humboldt Gresgud and Ebenezer Schwartz, both workers at the CFAISI. The initial swathe of infection did, it is true, radiate from a point coinciding with the location of the institute, and the laboratory where Doctors Gresgud and Schwartz pursued their research is situated on the tenth floor of the building, an ideal height from which to spread a plume of windblown spores over the cotton plantations east of Rootesville.



A comprehensive history of the disease should mention that Doctors Gresgud and Schwartz were allegedly infamous for their unbending misogyny. It should also be stated that they both vehemently denied being responsible for the epidemic, claiming that they merely identified the cause of the disease where other medical laboratories had failed. Both refused to voice any sympathy for the epidemic's victims.

Symptoms are insidious at onset, but progressive. The early phase resembles Tourette's syndrome, the patient being unable to refrain from outbursts of extravagant language. Male sufferers first mutter then shout with increasing frequency: "Fie, fie, fickle female!" Female victims in similar manner splutter: "Bastards! Bastards! Bastards!" The preliminary phase typically lasts three days. A vehement revulsion for the opposite sex then sets in, both psychologically and physically, developing in not less than 48 hours into a dread which equals the intensity of rabid hydrophobia.

Once the early phases are gone through sufferers are able to function normally provided they meet with no visual, auditory, olfactory or verbal stimuli reminding them of the opposite sex.

The disorder has defied attempts at treatment. Segregation of sufferers is the only recourse. The rapid spread of the disease is surprising. It can be transmitted by skin contact, sneezing or even by breathing heavily in the same room. It also comes as a surprise that the sexual link binding the male and female sides of the human race can be so drastically severed. New forms of social organization have perforce arisen in those districts where the epidemic has taken hold, leading to the emergence of a new study in the field of sociology.

For all that, it is necessary to interject a professional note. Criticism must be levelled at those who step outside medical bounds and spread alarmist prognoses, stating that should it come about that males and females no longer reproduce in partnership they are henceforth, by definition, separate species; predicting the division of the planet into male and female hemispheres, and so forth. It is admittedly fortuitous that non-sexual methods of reproduction are now available. But it should be stressed that medical science has faced greater challenges than simple gender phobia before now. If formic gender disorder introduces any novelty into the medical record it is that this is the first contagious disease in history whose sufferers unanimously proclaim it an inestimable boon to humankind.

References

- Baker, C. N. M. Bacteria and fungal spores in neural tissues of east-of-Rootesville field rodents. *Quarterly Journal of Immunology* (Spring 2030).
- Department of Neuropathology, Fremmen Hospital, San Francisco. Cellular examination of formic gender disorder cadaver. *Quarterly Journal of Immunology* (fall 2030).
- Department of Neuropathology, Fremmen Hospital, San Francisco. Aberrant dendritic growth and its agents. *Neurological Science Review* (Spring 2031).

Submitted by: John Barrington, MD. ■
Barrington J. Bayley's science-fiction production spans half a century. The novels *The Soul of the Robot*, *The Fall of Chronopolis* and *The Zen Gun* are typical of his work.

JACEY